2019

# Characteristic time courses of electrocorticographic signals during speech

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Dissertation

# CHARACTERISTIC TIME COURSES OF ELECTROCORTICOGRAPHIC SIGNALS DURING SPEECH

by

**SCOTT ANDREW KUZDEBA**

B.S., Syracuse University, 2010
M.S., Worcester Polytechnic Institute, 2012

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2019

Approved by

First Reader _____

Frank Guenther, Ph.D.
Professor, Department of Speech Language and Hearing Sciences

Second Reader _____

Mark Kramer, Ph.D.
Associate Professor, Department of Mathematics and Statistics

Third Reader _____

Brandon Hombs, Ph.D.
Chief Technology Officer, Eigen, LLC

## DEDICATION

For my daughters, Anya and Nadia, who I had the pleasure of watching gain the gift of speech, and to all those who cannot fully exercise that gift, may we continue to gain knowledge and break down barriers so that one day they may communicate freely.

## ACKNOWLEDGMENTS

A work of this magnitude and duration always consists of contributions from many individuals along the way, and this work is no different. The contributions themselves come in many forms, all of which have been integral to my success.

First and foremost, I must acknowledge my wife, Hillary, who has supported me through the ups and downs of this journey. She helped to keep me pointed in the right direction, lifted me up when I was down, and allowed me to maintain my perspective through these years. She has also taken on more than her fair share in our relationship so I could focus on completing this work, including raising our two daughters, caring for an ever growing menagerie of animals, and managing two separate moves. A sincere thank you and debt of appreciation Hillary.

I would also like to recognize my colleagues at BAE Systems who gave me the flexibility and support structure to continue to work full time while completing my PhD. A special thanks to BU alums Brad Rhodes and Majid Zandipour for steering me towards the computational neuroscience program at Boston University and providing the initial guidance I needed to start the program. A big thanks to Josh Niedzwiecki, John Tranquilli, Dave Logan, John Hogan, and Bill Watson for your support and vested interest throughout.

Along the way I also received support from many others that are too numerous to account for here. I would like to extend a special thanks to my parents, Judy and Peter, and in-laws, Jean and Steve, for helping throughout the years, especially Jean who has spent many nights at our home helping us keep our heads above water during this busy

time. To the rest of my family members who have also provided their love and support along the way, thank you.

Next I turn to those who collaborated and provided guidance to the contributions of this work. First and foremost, I owe my advisor and mentor Frank Guenther a huge thanks. He always kept me engaged, interested in learning and discovering more, and continually helped me to extract my best work. Further, he provided the flexibility and understanding I needed in order to juggle family and work during my studies. I could not have finished this work without his consistent (and very, very patient) support.

Jeremy Greenlee of the University of Iowa, not only collaborated to provide the data used in this work, but also opened his doors to me, allowing me to visit his lab and witness a brain surgery in the process. He also provided great feedback on my publications and was more than willing to provide advice and suggestions to a student he barely knew.

I am also extremely grateful to the rest of my committee for their guidance, interest, and investment throughout the years. Ayoub Daliri helped me get setup with my initial preprocessing pipeline and was my "go to" for questions about electrocorticography. I had many great methods conversations with Brandon Hombs and Mark Kramer, both of whom helped my interests grow in all things computational. Finally, Jason Bohland always brought an independent, objective viewpoint and was able to help me step back from my work to see it as others would.

I also owe a huge debt of gratitude to Jason Tourville. Jason was instrumental in helping me understand the anatomy of the brain and the underlying physiology of the

processes I was studying. He also assisted me with interpreting various results and was always responsive. I couldn't have completed the work without his help. Alfonso Nieto-Castañón was also always willing to provide advice and suggestions related to methods to try, and I greatly benefited from his thought provoking ideas.

The BU Guenther Speech Lab provided me with an extremely valuable, collaborative and interactive environment to learn and develop within. I am specifically grateful for the interactions I had with Emily Stephen and Byron Galbraith when I was a junior member of the lab. Thanks to the current and past members as well: Andrés Salazar-Gómez, Nan Jia, Spencer Torene, Jennifer Segawa, Mikhail Panko, Dante-Smith, Saul Frankford, Matthias Heyne, Megan Thompson, Elaine Kearney, Liz Heller Murray, and Matthew Masapollo. In addition, I am extremely appreciative of everything that Bobbie Holland does for the lab.

The BU Graduate Program for Neuroscience is a unique community that I have greatly benefited from during my time at BU under Shelley Russek's guidance. I also greatly benefited from my interactions with Sandi Grasso, who was instrumental in guiding me through the program. Many thanks to both of you. There are many others that are too numerous to name here that I am also thankful for. It truly was journey that I had the pleasure of sharing with so many.

Lastly, I am humbled by those who are undergoing stressful clinical procedures and volunteer their time, energy, and bodies as research participants. Our research participants are the true stars in helping to move the science forward and I am forever grateful for their generosity. Thank you.

**CHARACTERISTIC TIME COURSES OF ELECTROCORTICOGRAPHIC SIGNALS DURING SPEECH**

**SCOTT ANDREW KUZDEBA**

Boston University School of Medicine, 2019

Major Professor: Frank Guenther, Ph.D., Professor, Department of Speech, Language and Hearing Sciences

ABSTRACT

Electrophysiology has produced a wealth of information concerning characteristic patterns of neural activity underlying movement control in non-human primates. Such patterns differentiate functional classes of neurons and illuminate neural computations underlying different stages of motor planning and execution. The scarcity of high-resolution electrophysiological recordings in humans has hindered such descriptions of brain activity during uniquely human acts such as speech production.

The goal of this dissertation was to identify and quantitatively characterize canonical temporal profiles of neural activity measured using surface and depth electrocorticography electrodes while pre-surgical epilepsy patients read aloud monosyllabic utterances. An unsupervised iterative clustering procedure was combined with a novel Kalman filter-based trend analysis to identify characteristic activity time courses that occurred across multiple subjects. A nonlinear distance measure was used to emphasize similarity at key portions of the activity profiles, including signal peaks. Eight canonical activity patterns were identified. These activity profiles fell broadly into two

classes: symmetric profiles in which activity rises and falls at approximately the same rate, and ramp profiles in which activity rises relatively quickly and falls off gradually. Distinct characteristic time courses were found during four different task stages: early processing of the orthographic stimulus, phonological-to-motor processing, motor execution, and auditory processing of self-produced speech, with activity offset ramps in earlier stages approximately matching activity onset rates in later stages. The addition of an anatomical constraint to the distance measure to encourage clusters to form within local brain regions did not significantly change results. The anatomically constrained results showed a further subdivision of the eight canonical activity patterns, with the subdivisions primarily stemming from sub-clusters that are anatomically distinct across different brain regions, but maintained the base activity pattern of their parent cluster from the analysis without the anatomically constrained distance measure. The analysis tools developed herein provide a powerful means for identifying and quantitatively characterizing the neural computations underlying human speech production and may apply to other cognitive and behavioral domains.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF EQUATIONS

## LIST OF ABBREVIATIONS

AC ..... Anatomically Constrained

AP ..... Auditory Processing

BCI ..... Brain Computer Interface

cNMF ..... convex Non-negative Matrix Factorization

cs ..... central sulcus

dB ..... decibels

DIVA ..... Directions Into Velocities of Articulators (model)

ECoG ..... Electrocorticography

EEG ..... Electroencephalography

ESP ..... Early Stimulus Processing

fMRI ..... functional Magnetic Resonance Imaging

HG ..... Heschl's Gyrus

Hz ..... Hertz

iEEG ..... Intracranial Electroencephalography

IFG ..... Inferior Frontal Gyrus

Ins ..... Insula

ME ..... Motor Execution

MFg ..... Middle Frontal gyrus

MTF ..... Modulation Transfer Function

MTG ..... Middle Temporal Gyrus

ms ..... milliseconds

np.....normalized power

Occ.....Occipital

pAC.....posterior Auditory Cortex

PCA..... Principal Component Analysis

PMC.....Premotor Cortex

PP..... Planum Polare

PT.....Planum Temporale

PTC.....Posterior Temporal Cortex

PtM..... Phonological-to-Motor Processing

RBF..... Radial Basis Function

SC..... Somatosensory Cortex

sf.....Sylvian fissure

SMA.....Supplementary Motor Area

SMG..... Supramarginal Gyrus

STG..... Superior Temporal Gyrus

STRF.....Spectrotemporal Receptive Field

vMC.....ventral Motor Cortex

vPMC..... ventral Promotor Cortex

vSC.....ventral Somatosensory Cortex

# CHAPTER I: INTRODUCTION

## I.1 Motivation

Scientific discovery has trended from a reliance on observations to data analysis (Hey Tansley et al., 2009). The current rate of data discovery is modulated in a large part by the data available, which in turn is a function of the recording methods and the analysis methods, amongst many other things. Neuroscience is no exception. Early studies relied on behavioral observations. Slowly over time the data recording methods have become more sophisticated, allowing for recordings down at the single electrode level. The construction of new analysis methods follows the development and implementation of new recording methods. Together, the back-and-forth interaction of gathering new types of data and the new analyses to understand the data have continued to shed new insights into how the brain function.

Electrophysiological recording methods have greatly accelerated the pace of discovery throughout the 20$^{th}$ and 21$^{st}$ century, with many sub-fields of neuroscience turning to animal models to gain insights, i.e., memory (O'Keefe and Dostrovsky, 1971) and vision (Hubel and Wiesel, 1959). Complex functions that are unique to humans, i.e., language, have enjoyed some of this accelerated rate of discovery, but have lacked the gains made in other, not human specific functions. The primary reason for this lag has been the limited ability to collect large quantities of different types of data from human subjects, and to do so in invasive ways, i.e., electrodes on the surface of the brain.

Electrocorticography (ECoG), or intracranial electroencephalography (iEEG), has begun to become a mainstay in clinical evaluation of patients undergoing surgery,

providing a research opportunity for high temporal resolution data to be collected from awake subjects performing tasks. ECoG primarily uses electrodes placed on the surface of the brain, i.e. surface electrodes, but is seeing an emergence of depth electrode for their clinical benefits. While the basis of ECoG recordings is not new, its clinical use, and use in research experiments, in awake human patients is recently becoming more widespread with the potential to provide new discoveries.

This data recording method has allowed for new datasets for language, and in particular speech production, to be collected. This new type of data, in turn, is motivating the construction of new analysis methods. The work of this thesis heeds this motivation and joins the scientific discovery cycle by developing new analysis methods for analyzing ECoG during a speech production task. This helps to provide new insights and discoveries to a relatively new data collection method utilized in human clinical settings.

In particular, this work adds two novel forms of analysis to our investigation of speech as captured by ECoG. We pose our approach as an unsupervised and data-driven discovery of the underlying canonical neural activity during speech production. In doing so, we add to an emerging field of ECoG functional clustering analysis (i.e., Hamilton et al., 2018; Leonard et al., 2019). First, we develop a novel contribution to the methodology for discovering canonical activity through the use clustering analysis. In particular, we add a nonlinear distance measure to assess similarity between individual activity patterns, allowing emphasis to be put on the important parts of the activity, such as the peak. Second, we develop a structured framework to quantitatively characterizing

the discovered activity patterns. We do so by learning the underlying trend in the activity using a Kalman filter and detect when activity no longer fits the trend and there is a discrete change to a new trend. The combination of these two analysis methods enables us to gain new insights into neural processing of speech production, and, more importantly, sets us down a new analysis path that allows us to ask new types of questions to unlock more of the secrets of how the brain processes speech.

A brief discussion of speech follows, along with a focused discussion of machine learning and some of the insights that lead to the development of our methodology. This chapter then concludes with a summary of this dissertation.

## I.2 Speech

Speech, and more generally language, is one of the most complex functions that humans perform. It is also one of the functions that help to set them apart from most other species. Speech plays a critical role in our everyday life and has helped to build civilizations and social constructs that informally define how we live our lives and interact with others. With such as central role in our lives, speech has the power to enable individuals to accomplish great things. This can be witnessed by the progression of human history, which has largely been driven by society becoming more interconnected with language being the primary connecting factor.

However, speech also has the power to isolate individuals who suffer from aphasias, other disorders affecting language and speech, or other limiting conditions. Understanding how the brain processes speech has many benefits, including the potential

to provide these individuals who struggle with speech, or cannot speak at all, the ability to speak.

Lots of knowledge has been gained about speech through behavioral studies (from both normal and abnormal functioning brains), neuroimaging studies (i.e., functional magnetic resonance imaging, fMRI), electrical recordings (i.e., electroencephalogram, EEG), and, in rare cases, local field potentials (e.g., refer to (Stevens, 2000) for an account of how these methods contribute to our understanding of speech sounds). These studies have combined to provide coherent models for how speech production is conducted (e.g., refer to (Guenther, 2016) for a detailed model of speech motor control), with additional studies attempting to tie together an overarching theory of language (Friederici, 2017). This collection of methods, however, typically trade off temporal resolution with spatial localization. Meta and cross-study analyses attempt to pull together the spatial results gained from one method with the temporal results of another, but results from this can only go so far and often become too smoothed, or averaged, to provide enough detail to fully characterize the time courses of speech.

Electrocorticography (ECoG), both surface and depth electrodes, provides a unique source for generating fine temporal resolution while still providing good spatial resolution (Ray et al., 2008; Miller et al., 2009). Speech is on the order of hundreds of milliseconds, well within the recording resolution of ECoG. Thus, the impact that ECoG can have on our ability to understand speech is immense. Emerging research in this area has started to help bridge the gap between prior studies with varying degrees of temporal

and spatial resolution, providing more detailed temporal understandings of neural processing during speech production.

In this work we add to the emerging field of ECoG speech research. We analyze ECoG data collected while subjects read aloud monosyllabic utterances. We make explicit use of the unique resolution (temporal and spatial) that ECoG can capture, where the temporal profiles of the ECoG recordings capture underlying neural activity. We discover from these recordings a distinct set of canonical activity patterns that are a characteristic set for speech production, broken across four processing stages: early processing of the orthographic stimulus, phonological-to-motor processing, motor execution, and auditory processing of self-produced speech.

### I.3 Machine Learning

The rate of discovery is accelerating across many fields because of the shift away from observations to the analysis of data. Further, many of these studies are shifting away from theory and explicit modeling towards data driven methods. Many of these data driven methods are based on machine learning, where algorithms are constructed to process the data and learn the underlying trends and representations that are naturally present in the data. This contrasts with theory or model-based approaches where the trends or representations are described by an expert in the field and explicitly built into the algorithm and data is typically only used to fit parameters of those models. We make use of this data driven approach in this dissertation.

Unsupervised machine learning is a particular data driven approach where the data is unlabeled and the algorithm(s) find structure that exist within the data in the absence of any additional details about the data. This is opposed to supervised machine learning, where labels are provided with the data and are used in the training process to provide guidance, or supervision, for the data structure that the algorithm finds. For example, natural language processing is a supervised learning task where labels, in the form of words, would be provided during training to annotate an acoustic signal (Jurafsky and Martin, 2009). The algorithm would use this to learn the structure within the acoustic signal that can classify words, text, from an acoustic signal. Then during testing the algorithm would predict what words are present in an unlabeled acoustic signal. Unsupervised methods would not have the labels present, so would not necessarily break up the acoustic signal by word segmentations. Instead the acoustic signal would be broken up by some other measure, typically the degree of similarity that exists between different parts of the acoustic signal, resulting in a segmentation by naturally occurring patterns within the acoustic signal, which may or may not be grounded in words.

In this work, we will take an unsupervised learning approach that discovers structure that is naturally present in the ECoG data, as a surrogate for neural processing. In doing so, we are not trying to guide the representation in the form of labels, but by allowing it to group the ECoG signals based on their degree of similarity through clustering analysis. We preprocess the data to time segments of speech production for each electrode and then let the unsupervised learning determine how to group, or cluster, the individual segments together. This results in characteristic time courses, or activity

patterns, that are present during speech production. This approach does not place requirements on how the grouping is conducted in any way. Thus, the results are not guided to be broken up anatomically, functionally, or in any other way unless the underlying data has those break-downs naturally existing within it.

Data driven methods have also become more common in their use for describing results. Model-based methods fit parameters of a model with the collected data and thus have an explicit framework in which to describe the outcome, with the parameters that are fit typically capable of describing important details of the model or theory. Having an explicit model form allows for the model to be constructed to describe specific parts of the underlying data, but does not provide the flexibility to extract from the data what should be described. Using a model-free approach, on the other hand, allows for the underlying nature of the data to dictate what should be described, but often suffers from interpretability (Lipton, 2016; Turner, 2016; Geffner, 2018).

In this work, we will use something in-between these two data driven approaches. We functionally construct a model class, a linear Kalman filter, to describe the trends present in the activity patterns. To enable the underlying nature of the data to dictate where the trends change and how they should be represented, we add in an adaptive learning rate and change point detection that allows for the current trend models to be stopped and new models to be fit if the naturally present trend in the data changes such that the current model no longer represents the data. The outcome of this approach is a data driven way of getting quantitative measures to describe the activity patterns, where a

model-based framework is used to guide the quantitative measures to be descriptive for the activity trends, overcoming challenges in interpretability.

Additional constraints and considerations, such as anatomical similarity, are also used in various parts of this work, but not discussed here. These additions will be covered in detail in their respective locations, or relegated to an appendix. The unsupervised clustering approach and the Kalman filter-based change point detection are the two novel parts of our methodology that enable the findings that will be presented in this work and provide a new data driven analysis approach to understanding speech production from ECoG.

### I.4 Summary of Dissertation

The goal of this research was to identify the characteristic time courses of neural processing during speech production. Electrocorticographic recordings were used to measure neural activity within and on the surface of the brain. We use these recordings to discover the set of characteristic time courses that are naturally present during speech production and use novel methods to quantitatively describe the characterizations. First, in CHAPTER II, we review the prior art and background in this area of research and situate the work of this dissertation amongst this body of work. Next, in CHAPTER III we lay out the methods that we use for our analysis, including a set of novel additions to the analysis of speech from ECoG. CHAPTER III also presents the results of these methods and discusses the findings. The bulk of the work of this dissertation is contained within CHAPTER III. CHAPTER IV builds on CHAPTER III, relooking at what

happens to the findings if anatomical constraints are placed on the analysis to encourage resulting groupings to be locally clustered within the brain. Finally, CHAPTER V lays out future directions of this line of research. A set of appendices are included to provide additional details, motivation, and alternatives for the methods and results that are not covered in detail within the chapters.

Taken together, this collection of work lays out a new way to analyze speech production from ECoG, which itself is an emerging method, and a framework for quantitatively characterizing activity patterns that are found within the ECoG signals. The rest of this section provides a brief summary of each chapter.

### *I.4.i Chapter II: Background*

CHAPTER II provides a survey of the background for this research area. It is laid out to provide a loose pedigree of the line of research that the current work builds on, provides results from similar studies, motivates the need for this work, and situates this work amongst the prior art.

The chapter starts off with a discussion on speech, quickly covering the early work in the field, such as the discovery of Broca's and Wernicke's area (Broca, 1861; Wernicke, 1874), and building to current models of speech, such as those laid out in (Guenther, 2016). Much of the prior work covered here is from behavioral and neuroimaging studies, both of which added significant understanding to the field. However, these types of studies are limited in their temporal resolution, and are not able to accurately capture neural processing at the same rate of speech, which is on the order of a hundred milliseconds.

Electrocorticography (ECoG) is then covered next. ECoG is the recording method used for this dissertation and supplements other recording methods by providing high temporal resolution for a sampling rate that can accurately capture speech (Ojemann et al., 2013). This section of the chapter walks through the early work on ECoG, laying out the properties of the recordings that will be used in the later chapters, such as the important finding that ECoG high gamma power captures a similar effect as local field potentials (Ray et al., 2008). Additional details on some of the early findings and properties of ECoG analysis are discussed. This is followed by a discussion on a group of studies we refer to as *first order analyses*. These studies begin to explore questions of neural processing of speech by asking *if* activity changes during a speech task, but do not explore *how* activity might be changing. In particular, many of these studies look at if there is a statistically significant change in high gamma ECoG power during speech production, such as the organization of speech articulators and phonetic features within the sensorimotor cortex (Chang et al., 2010; Bouchard et al., 2013). Next, a set of *second order analysis* studies are reviewed. This set of research differs from first order analyses in that the questions start to look at *how* activity is being modulated during a task. The methods used here heavily rely on prior knowledge and theory to set the approach, often resorting to model-based approaches that capture the temporal nature of the signal, but are still limited by the priors that are built into the system, such as (Ozker et al., 2017) who looked at the differences in the posterior superior temporal gyrus to speech and non-speech sounds using linear mixed effects models. Some work here has started to turn to machine learning, such as (Lotte et al., 2015) who used machine learning to derive the

features to feed into their model, but the methods within this grouping of studies still relies heavily on prior knowledge built-in to multiple parts of the methodology.

This led to an emerging area of research that relies on a class of unsupervised machine learning methods: clustering. Cluster analysis finds groupings in data given some measure of similarity. Few studies have ventured into this method for ECoG analysis, with the primary group looking at this line of research being the Edward Chang lab at the University of California San Francisco. Several studies utilize clustering, but do not delve into the exploration of the characteristic activity patterns. One study from the Chang lab, (Leonard et al., 2019), is the closest to the work of this dissertation and begins to look at characteristic activity patterns without a quantitative description. In Leonard et al., a soft clustering approach – convex non-negative matrix factorization (cNMF) – is used to find groups that break up the spatiotemporal dynamics of speech during a word repetition task. This approach has several limitations that the current work looks to overcome.

Next, we return to speech, briefly bringing everything full circle with a more general discussion of speech. We use this section to help motivate the continued use of ECoG for studying speech and adding to our current theories and models. We discuss the need for the new analysis methods, such as those that are developed in this work, to provide additional insights. In particular, we aim to motivate the need in providing a more detailed account of the temporal nature of neural processing during speech. We prompt this by looking at the gains that have been made recently with brain-computer

interfaces (Guenther et al., 2009; Brumberg et al., 2010), but note that the performance is still not yet at a level to make it useable.

Finally, we wrap up the chapter with a discussion where we situate the work of this dissertation amongst the prior art. We summarize the motivation for the use of electrocorticography to record data at a sampling rate that captures the temporal dynamics of speech. We also lay out the need for a data-driven methodology to discover the characteristic time courses without enforcing prior assumptions or knowledge. The use of clustering as a source of data-driven discovery is discussed, highlighting its ability to find characteristic activity patterns. Lastly, we highlight our novel additional contribution of a framework to quantitatively describe and compare the resulting characteristic activity patterns.

### *I.4.ii Chapter III: Canonical Speech Clusters*

CHAPTER III contains the bulk of the work of this dissertation and is setup to be a standalone section that can be read and understood on its own. This chapter begins with a significance statement for the research contained within. This is followed by an introduction, which is a more focused and concise version of CHAPTER II. The methods are laid out next, followed by the results and discussion. A supplementary section is also provided that gives some additional results supporting claims made earlier in this chapter. We now summarize the chapter.

Surface and depth electrocorticography (ECoG) electrodes collected data from five subjects who read aloud monosyllabic words. The high gamma (70 – 150 Hz) power temporal profile was taken for each electrode. Speech trails were then extracted from the

high gamma activity for each electrode and averaged across trials under two alignment conditions: aligned by the presentation of the orthographic stimulus and by the onset of voicing. Data was then corrected by a baseline, non-speech period, i.e., data was z-scored based on the non-speech period existing one second to half a second prior to the stimulus presentation. Only electrodes showing significant deviation from the baseline activity levels during the trial were used for further analysis.

This data is passed through a hybrid clustering algorithm, which uses a data driven approach to group electrodes together based on the similarity in their high gamma power activity patterns. To enforce similarity during important parts of the activity, i.e., activity peaks, a nonlinear distance measure is used. This is different than prior studies, which use linear measures and therefore put equal weight in slight deviations from baseline activity and deviations in peak activity.

Eight clusters were found from this method. These clusters make up a set of canonical activity patterns for speech production. Even though the clustering method is unsupervised and does not incorporate priors on the speech network, the results are broken up by the stages of speech production: early response to the orthographic stimulus, motor planning, motor execution, and the auditory feedback to self-generated speech. Each processing step is made up of two clusters.

To characterize the clusters, a new method was developed. Characterizing the clusters consists of quantitatively describing the activity pattern, both in terms of activity activation and decay rates, as well as determining when activity trends change. A Kalman filter is used to determine the trends present within the activity patterns. Kalman

filters have been used extensively to track and predict time series data and provide an attractive framework for discovering data trends. To enable the detection of points of discrete changes in the trends, we modify the Kalman filter to perform change detection. We do so by adding in an adaptive learning rate to the Kalman filter gain term, which functions to bias the Kalman filter towards its prediction step and away from its update step as time progresses. This has the effect of allowing the data to determine the parameterization of the filter as it is being initialized, i.e., learning the naturally present trend, but over time relying on the filter prediction and less on the observed data, i.e., a shift to a more model-based representation of the current trend. A change is declared when the observed data statistically deviates too much from the learned model prediction. This provides a means to discretely segment an activity pattern and quantitatively determine the trends present.

The resulting clusters show activity patterns that are described as *symmetric*, where activity increase and decrease rates (trends) are approximately equal, or as *ramp*, where the increase rate (trend) of activity has a larger magnitude than the decrease. All processing steps exhibited both activity patterns, expect for motor execution which only showed symmetric activity. The symmetric activity within the motor execution group was subdivided into a broad and narrow activity pattern, based on the overall duration of the activity, as well as the extent of the activity plateau. The decrease in activity in one processing step roughly mirrors the increase in activity in the next processing step, hinting at a potential gradual handoff of neural processing between brain regions.

Electrodes responding to the stimulus presentation are found bilaterally in posterior and ventral temporal regions with some frontal lobe activity, largely following previous work that found these regions active during object identification (Chaumon et al., 2014), saccadic eye movements (Zhou and Shu, 2017), and reading of non-word orthographic stimuli (Vigneau et al., 2005). In the motor planning group, which we term phonological-to-motor processing, activity is largely limited to the left hemisphere, supporting prior findings of a left lateral bias to speech motor planning (Guenther et al., 2006; Guenther, 2016). Motor execution is found bilaterally with widespread activity. A substantial percentage of motor execution electrodes are found surrounding the central sulcus in the ventral sensorimotor cortex, where motor and somatosensory representations of speech articulators are found (Penfield and Roberts, 1959; Takai et al., 2010; Bouchard et al., 2013; Guenther, 2016). Finally, auditory processing electrodes are found primarily within the auditory cortex.

### *I.4.iii Chapter IV: Anatomically Constrained Clusters*

CHAPTER IV builds on CHAPTER III, utilizing the same methodology with one change. The focus of this chapter is to see what happens when anatomical constraints are placed on the clusters that the electrodes can form. Many prior studies localize specific functions to distinct regions of the brain, and speech is no different (i.e., Broca et al., 1861; Broca, 1865; Wernicke, 1874). Therefore, this study looks to see if characteristic time courses exist within local areas of the brain and, if so, how they compare to the results found without the constraint added in, as covered in CHAPTER III.

The anatomical constraint is added as a modification of to the distance measure used for clustering. The distance measure in CHAPTER III is a nonlinear measure of the similarity in the time courses of different electrodes' activity patterns, which we refer to as the temporal difference. The anatomical constraint is added as a weighted mixing term to this temporal distance measure, adding an anatomical similarity component to the distance measure. To assess anatomical similarity, a radial basis function, similar to the Gaussian kernel in (Kubanek and Schalk, 2015), is used on the Montreal neurological institute (MNI) coordinates of the electrodes. This places more weight on nearby electrodes and less on far away electrodes. To allow for bilateral clusters, the absolute value of the lateral MNI component is taken.

A larger set of resulting clusters stem from this, as compared to CHAPTER III. The resulting characteristic time courses (clusters) generally subdivide into the eight canonical time courses found in CHAPTER III. The resulting clusters of this analysis are also separated by the same four processing stages and two activity shapes, symmetric and ramp. The characteristic time courses here, however, show further subdivisions, breaking up individual time courses from CHAPTER III into multiple clusters that are divisible primarily by the brain region that they cover. For example, the phonological-to-motor ramp temporal profile is separated into four clusters in this analysis: one in ventral temporal and occipital regions, one localized to the frontal lobe, one surrounding the ventral portion of the central sulcus, and the final one more diffuse, covering sensorimotor, somatosensory, and auditory areas.

Some minor differences in activity patterns did emerge. For example, in the phonological-to-motor ramp group just described, one of the clusters shows a symmetric shape for most of its activity pattern, but during its decay it transitions into a long slow decay pattern, more representative of the ramp pattern. In the motor execution group, two new activity patterns are seen. One pattern is an inverse ramp, showing a slow rise in activity levels and then a quick decay. This is the inverse of the ramp shapes that we have seen previously, but aligns with nonhuman single unit motor cortex recordings (Cheney and Fetz, 1980). We refer to the other new activity pattern as the suppressed narrow symmetric activity pattern. This cluster shows the activity of the motor execution symmetric narrow group, but has a brief period before activity onset where high gamma power is suppressed below baseline activity levels occurring just after stimulus presentation, hence the inclusion of suppression in its naming.

In conclusion, some differences do exist between the anatomically constrained (CHAPTER IV) and non-constrained (CHAPTER III) analyses. However, where there are differences, the supporting number of electrodes in the clusters of the constrained analyses is less than the number of electrodes in other clusters, putting less confidence in the results from these clusters. For example, one of the new activity shapes just described only has two electrodes in its cluster. There is a large overlap in the findings across the two analyses between clusters that have more electrodes, with the only differences being a subdivision of clusters in the analysis of this chapter by anatomical boundaries. Therefore, the findings of this chapter are found to be largely in support of the findings of CHAPTER III.

### *I.4.iv Chapter V: Future Directions*

Finally, CHAPTER V provides concluding thoughts in the form of a future roadmap. It is our hope that others will continue with this line of research. In this chapter we lay out the directions that we believe could be taken, both in the near term and further down the road. In doing so, we take a narrow view on the next research steps that naturally follows our work presented herein and also provide implications for broader impacts outside the specific focus of this work.

### *I.4.v Appendices*

Additionally, we provide a set of appendices that provide more details on the methods, motivations, and results. The majority of these appendices are in support of the work done in CHAPTER III. We provide a brief summary of each appendix.

APPENDIX A provides more details on the Kalman filter-based change point detection method used for the quantitative characterization of the canonical activity patterns. It provides motivation for the design choices and discusses the alternatives that were considered.

APPENDIX B provides more details on clustering, which played a central role in the work of this dissertation. It loosely follows a similar structure as APPENDIX A, providing more detail on the method, motivation for the design choices, and discussion on alternatives. Many options exist for clustering. This appendix helps navigate these choices and also put the current work in the context of what others have done. Several key areas discussed are the choice of a distance measure, the type of preprocessing conducted prior to clustering, determining how many clusters should be used, and how

electrodes are assigned to clusters, i.e., hard versus soft clustering. This section provides many additional details that will help navigate prior research and better understand the contributions of this work amongst them.

APPENDIX C provides more general details about the methods used, covering all aspects of the methods more broadly than the narrow focus of APPENDIX A and APPENDIX B or concise descriptions in CHAPTER III. The focus, again, is on providing additional details, motivation, and a discussion of alternatives. Particular focus is devoted to the choice of filtering.

Lastly, APPENDIX D provides some additional results that are not covered in detail elsewhere. CHAPTER III presents cluster results that exist across multiple subjects. In this appendix, we present the clusters that are left out of CHAPTER III since they only exist within a single subject. This set of clusters that each only come from one subject is mostly limited to one or a few electrodes. One cluster, however, has over 25 electrodes in it, coming from one subject and showing high gamma power suppression during the speech production task. Additionally, this appendix presents results when the analysis of CHAPTER III is rerun with only the surface electrodes, i.e., the depth electrodes are dropped. The results of this analysis have a very high overlap with the analysis that uses all of the electrodes (CHAPTER III), further supporting the claim that canonical activity patterns are found. The focus of CHAPTER III and CHAPTER IV is on high gamma power, but it is noted in CHAPTER II that other frequency bands have been found to also play a role in the neural processing of a task. The last part of this appendix presents results for one of these other frequency bands, beta (15 – 30 Hz),

which has been found to show a suppression in local activity during a task (Pfurtscheller and Lopes da Silva, 1999), amongst other things.

# CHAPTER II: Background

## II.1 Speech

Our understanding of speech has evolved as we have invented and utilized new methods for recording and analyzing brain activity during speech production and perception. Initial high level understanding was gained through behavioral, lesion, and aphasia studies, i.e., (Broca, 1861; Wernicke, 1874). A more detailed understanding of speech followed with the advent of neuroimaging and electrophysiology recording methods, such as functional magnetic resonance imaging (fMRI). Findings from these methods have led to current knowledge of the human language cortex widely distributed over large-scale networks in the temporal, parietal, and frontal lobes (Ojemann et al., 1989; Indefrey and Levelt, 2000, 2004; Hickok and Poeppel, 2004). This has brought about language theory postulating that neural responses to speech exist throughout the brain, tying together low-level auditory and motor processing with higher-level language features from phonemes to phrases (Giraud and Poeppel, 2012; Hagoort and Indefrey, 2014; Ding et al., 2016).

The combined use of functional imaging, electrophysiological recording, and detailed behavioral, lesion, and aphasia studies has allowed for the study of various components of speech, providing detail about both individual brain regions and functional connectivity. A network of key language processing regions has been found. This network is widespread involving the temporal, parietal, and frontal lobes, including posterior auditory cortex (pAC), ventral premotor cortex (vPMC), ventral motor cortex (vMC), ventral somatosensory cortex (vSC), and supplementary motor area (SMA), to

name a few (Turkeltaub et al., 2002; Guenther et al., 2006; Hickok and Poeppel, 2007; Ghosh et al., 2008; Friederici, 2012; Hagoort, 2013). This network has been analyzed and teased apart through many studies involving subjects verbally responding to acoustic or visual stimuli, confirming and providing updates to aspects of the speech network with much of the work focused on perisylvian regions (Price et al., 1996; Warburton et al., 1996; Vallar et al., 1997; Anderson et al., 1999; Baldo et al., 2008; Pei et al., 2011b; Herman et al., 2013; Majerus, 2013; Moritz-Gasser and Duffau, 2013; Hope et al., 2014; Parker Jones et al., 2014). Additional work has extended out to include multisensory speech perception (Magnotti and Beauchamp, 2017), but is beyond the scope of this dissertation.

We use an example of a speech task to illustrate how individual studies and different methods contribute to our collective knowledge about the speech network. What follows is a brief walkthrough of the speech production processing steps from results in a set of studies where a subject repeats acoustic word stimuli. First, auditory processing of the stimulus is conducted within the superior temporal gyrus (STG) as frequency and temporal features of the acoustic stimulus are processed (Mesgarani et al., 2014), with distinct regions showing differential preference for low-level acoustic features and higher order phonetic features (Chang et al., 2010). Working memory is then invoked to maintaining the phonological sequences, involving inferior frontal and posterior regions spanning temporal and parietal cortices (Paulesu et al., 1993; Buchsbaum et al., 2005). Next, motor planning and articulation of the stimulus is performed in the prefrontal and ventral sensorimotor cortices (Bouchard et al., 2013).

Lastly, auditory feedback is processed to modulate articulation (Houde and Jordan, 1998; Hickok et al., 2011; Chang et al., 2013). Refer to (Guenther et al., 2006) for a full account of pulling this all together.

Modern methods and studies, like those laid out in the preceding paragraph, have confirmed, added to, or modified theories on how the brain processes speech. Current theories have used these findings from more recent techniques to build upon and replace earlier theories based on understanding gained from more limited methods and analyses (Lichtheim, 1885; Geschwind, 1979). Recent studies support leading theories of speech processing, both perception and production, having a dispersed functional flow across multiple brain areas, e.g., (Guenther, 2006). They also help to support higher level language models, e.g., a dual stream theory of language processing (Rauschecker, 1998; Hickok and Poeppel, 2004), where there is still many unanswered questions and unknown aspects and are out of the scope of this dissertation.

These theories have led to speech models. Current speech models have built on older theories and models, collecting, aggregating, and fusing research from a large number of studies utilizing diverse techniques, tasks, and analyses. In this work, we will utilize the speech network described in the Directions Into Velocities of Articulators (DIVA) model (see (Guenther et al., 2006; Guenther, 2016) for a review) to position our findings within the speech network. The DIVA model has been extensively analyzed against the findings of prior studies, including a large body of fMRI analyses (e.g., refer to (Bohland and Guenther, 2006; Guenther et al., 2006)). This extensive validation has led to the model containing some of the leading theories of the speech network

(Guenther, 2016), including speech production (Guenther et al., 2015), which is the focus of this work.

## II.2 Electrocorticography

Electrocorticography (ECoG) is a relative newcomer to the analysis of speech processing, but it possesses particular benefits compared to some other methods – primarily its good spatial resolution and excellent temporal resolution (Ojemann et al., 2013). This excellent temporal resolution has allowed for new questions to be asked and has led to learning new things about how the brain processes speech in the various processing stages. ECoG activity have been connected directly to local field potentials, with ECoG high-gamma frequency (60 – 200 Hz) oscillations showing similar effects as local field potentials (Ray et al., 2008). ECoG has also been shown to closely correlated with fMRI blood-oxygenation-level-dependent (BOLD) responses (Lachaux et al., 2007; Hermes et al., 2012). These are powerful findings, allowing ECoG analyses to correlate with other neurophysiological recordings that have been more established in the field.

ECoG has shown the potential for both clinical and research applications. Clinically, ECoG has shown the capability to replace methods of mapping speech regions (Arya et al., 2015). From a research perspective, ECoG has shown the potential to provide detailed spatiotemporal relationships, such as early work showing the primary auditory cortex and STG are organized tonotopically (Wessinger et al., 1997; Bilecen et al., 1998; Formisano et al., 2003) and modulated by various spectral and temporal properties of sound (Giraud et al., 2000; Joris et al., 2004; Altmann et al., 2010; Leaver

and Rauschecker, 2010). Cortical evoked responses to spoken sentences have been demonstrated to be robust and selective to phonetic features over time periods extending out to 18 months (Rao et al., 2017). These findings showing ECoG signals maintaining stability over long periods of time contribute to validate that ECoG signals are capturing general aspects of speech and not just specific conditions during an experiment session. This helps build the case for ECoG's use in studying and understanding the human speech network. Thus, ECoG is a powerful tool to add to our research repertoire to aid in our on-going determination to unlock the secrets of neural processing of speech.

### *II.2.i Early Electrocorticography*

The properties of ECoG drove early research towards developing a more refined spatiotemporal understandings of speech neural processing. ECoG research initially focused on summary, high-level spatiotemporal ECoG analyses to map speech regions of the brain for clinical applications, such as for surgery (Crone et al., 2001b, 2001a, 2006; Towle et al., 2008). Research then shifted to understanding speech with improving methods, including a focus on spatiotemporal representations, to enable better use of the spatiotemporal resolution that ECoG provides (Chang et al., 2010; Edwards et al., 2010; Pei et al., 2011b).

Findings from early work with ECoG set the ground work for the development of new methods, including the following important findings. ECoG signals were found to follow a power law scaling with broadband amplitude changes directly indicating neural activity (Miller et al., 2009) and broadband power in the motor cortex predictive of task performance (Miller et al., 2012). An important ECoG frequency band, high gamma, was

shown to have power level phase-locked to theta oscillations (Canolty et al., 2006), track the speech acoustic envelops in auditory cortex (Kubanek et al., 2013), and exhibit suppression in auditory cortex during self-generated speech (Flinker et al., 2010).

These early findings set the foundations for later studies. The findings of band specific responses correlating with different types of neural activity has turned out to be a key early finding. In particular, high gamma frequency band correlating with local neural activity (e.g. (Miller et al., 2009)) has been the core preprocessing step of many following studies, including this one, as it has provided a way to discuss results in terms of neural activity levels within local brain regions. Neural properties of ECoG recordings are still being discovered as research progresses. For example, a recent study found that larger high gamma frequency activity is observed for higher cognitive demands when performing tasks with a higher speech working memory load (Kambara et al., 2017). This highlights that we still have more method development to do to extract properties present within the data of this recording technique.

### *II.2.ii First Order Analyses*

Several general classes of studies followed the discovery of key ECoG properties. We break them up into what we term *first* and *second order analyses*, with this subsection covering first order analyses and §II.2.iii covering second order analyses. We distinguish first and second order analyses by the level of information that is used from ECoG. First order analyses use high level, or summary information, from the data. Second order analyses aim to utilize more information about the underlying structure of the neural activity in the data. In this way, first order analyses are largely centered

around methods utilizing statistical significance tests to determine *if* there was a change in neural activity, while second order analyses utilize more complex methods to determine *how* there was a change in neural activity. The research conducted under this study focuses on second order analyses, but first we visit pertinent findings from first order analyses.

As mentioned, first order analyses are primarily concerned with using statistical significance tests to compare ECoG signals from stimuli or task periods to a baseline period to answer important questions about speech processing (e.g., see (Leuthardt et al., 2012; Brumberg et al., 2016)). Some questions explored and insights gained under this form of testing includes feedback control of vocal pitch (Chang et al., 2013), the organization of speech articulator and phonetic features within sensorimotor cortex (Chang et al., 2010; Bouchard et al., 2013), and differences from speaking versus listening (Cheung et al., 2016). Review (Martin et al., 2019) for a more complete review.

These comparative analyses are either between a baseline, typically non-speech period, and a task specific period or between different task specific states and are normally constrained to the power within a specific frequency band, such high gamma. The comparisons take the form of finding times of statistically significance difference between the comparison. These lower order analyses typically focus on some discrete aspect of speech that is already considered a feature of speech, including understanding the difference between vowels and consonants (Pei et al., 2011a), phonemes (Brumberg et al., 2011; Mugler et al., 2014), syllables (Blakely et al., 2008; Steinschneider et al., 2011), words (Kellis et al., 2010), and sentences (Zhang et al., 2012).

Studies on dynamics and functional networks followed suit, adding to our understanding of the interdependencies of the human speech network (Stephen et al., 2014; Stephen, 2015). Findings from these types of studies start to give insight into speech dynamics, getting closer to a better temporal accounting of neural activity. For example, high gamma task-evoked responses in Broca's area are constrained to the pre-articulation period (Flinker et al., 2015). Other studies have explored how context plays a role in speech, such as gestures (Mugler et al., 2018) and neighboring phonemes (Bouchard and Chang, 2014; Mugler et al., 2014), and how higher level goals, such as semantics and tasks, affect how information is processed specifically for speech (Mesgarani et al., 2014; Nourski et al., 2017).

This prior work has helped build a more detailed temporal understanding of speech, but has relied heavily on first order characterizations of ECoG activity, namely changes in average power levels, and has yet to uncover some of the higher order elements that may provide additional insight into human speech. Thus, they come up short in providing a detailed temporal profile of neural processing, which is one of the promises that ECoG enables. This body of work defines a spatiotemporal understanding of *if* cortical activity is present during a specific task, but it has not provided enough insight to begin to ask *how* activity is modulated during the task.

### *II.2.iii Second Order Analyses*

Task related changing dynamics have been found in the ECoG signals, such as reduced variability during stimulus onset (Dichter et al., 2016) and increased variance with increasing activity amplitudes (Tolhurst et al., 1983; Ma et al., 2006). These

changing dynamics hint at higher order aspects existing in ECoG signals, motivating the need for non-static analysis methods and second order analyses to understand how these dynamics manifest. Outside of speech, research in different brain regions and functions have shown the benefit of higher order analysis on characterizing neural activity, often turning to machine learning. For example, there is a structural hierarchical ordering of neurons in macaque inferior temporal cortex when presented visual object (refer to (Haxby et al., 2014) for further review on this example and others).

Machine learning has been one of the driving factors in moving from first order to second order analyses. The move has allowed methods to relax many of the assumptions and priors previously built into the models, enabling the ability to model dynamics and structure of ECoG signals that were not previously possible. Some limited focal areas within speech have started to use machine learning. For example, in (Lotte et al., 2015), machine learning was used to derive speech features, as opposed to being pre-specified during the model design process. Another area of focus has been to decode continuous speech directly from ECoG, borrowing techniques from similar fields, such as automatic speech recognition (Herff et al., 2015) or by using non-linear transforms (Pasley et al., 2012), with a recent influx of deep learning (Angrick et al., 2019; Anumanchipalli et al., 2019). While machine learning is just getting its foothold in speech, it has been used more widely in other neuroscience fields (see (Varoquaux and Thirion, 2014) for a neuroimaging review). This early work applying machine learning to speech has helped to confirm or add extra knowledge to our understanding of speech, but unfortunately still loses a lot of the important temporal information in the way they transform the ECoG

signals or suffers from interpretability. However, it should be noted that there is a growing trend in moving towards leveraging nonlinear functions. This trend is something that we will leverage in our methods.

Some work has been done to maintain the temporal nature of the signals. Maintaining the temporal characteristics of ECoG while still allowing for higher order structure has helped to confirm STG anatomically break-up of acoustic signal processing, with anterior STG showing greater neural activity to clear speech and posterior STG showing similar or greater activity when the input is replaced with speech-like noise (Ozker et al., 2017). This study used linear mixed effects and Bayesian models to understand the spatial and temporal dynamics of the ECoG signals as a function of the type of input signal. A different study looked at auditory phoneme blockage by adding noise to block a phoneme in the acoustic signal stimuli (Leonard et al., 2016). Auditory phoneme restoration was observed in the ECoG signals, with left frontal cortex seeing an increase in activity prior to auditory cortex restoration. This study used principal component analysis (PCA) and support vector machines (SVM) to model the ECoG signal in a representation that could maintain both the spatial and temporal dynamics while providing classification.

### II.3 Clustering

The spatiotemporal aspects of ECoG has enabled a more detailed dynamical model of neural activity in different brain regions, how it is modulated during speech, and how the different regions interact with one another. An emerging analysis method to

exploit the spatiotemporal aspects is clustering analysis. This unsupervised method is data-driven, and hence, has allowed for analysis with a larger set of parameters and less assumptions. This has further pushed us away from simple questions of *if* significant changes are observed and closer to gaining additional insight into *how*.

Similar to the research trend presented in §II.2, the focus within ECoG clustering studies involving speech have focused on the high gamma frequency band. In what follows, we will dive deeper into the studies that are most closely related to our work. We will begin by looking at studies that generally fall within this methodology and then will dive into two types of studies that are the closest related to our work, namely those involving spectrotemporal clustering and those that utilize convex non-negative matrix factorizations (cNMF).

### *II.3.i Speech Clustering Studies*

This first subsection dives into two speech clustering studies that more heavily utilizes priors and assumptions in their models. These priors and assumptions put the studies a little closer to what we called first order analyses, i.e., §II.2.ii, as their preprocessing removes some of the key temporal structure of the signals. They utilize a data-driven clustering approach, however, which aids in showing the promise in the potential of this method, while still capable of producing some key insights and findings.

First we will look at the work of (Berezutskaya et al., 2017). This work extends prior research that showed propagation of low-level acoustic features of speech from posterior STG to anterior STG by exploring what happens to neural activity next, past the STG, and how higher-level language processing areas, such as inferior frontal gyrus

(IFG), get involved. Their key finding was a propagation of temporal features of speech sounds (getting increasing coarse) along the ventral pathway of language processing.

Movie stimuli were encoded into a pre-defined 4D feature set consisting of spectral modulations, temporal modulations, frequency bins, and time points. Kernel ridge regression was used to model how these features estimate ECoG high gamma signals across 15 subjects. This model provided more accurate predictions in posterior STG for low-level speech encodings, with a gradient of prediction accuracy moving toward IFG, confirming earlier works, such as that by (Chang et al., 2010; Bouchard et al., 2013) as discussed in §II.2.ii. The regression coefficients found for each electrode were then fed into affinity propagation clustering (Frey and Dueck, 2007) to group the electrodes based on the similarity in regression coefficients. A Euclidean distance measure, i.e., linear, was used to measure similarity. Resulting clusters had the most variance along the temporal dimension, with clusters primarily separated by their activity time to the stimulus. Three clusters were located primarily within posterior STG, while the other three clusters primarily comprised of electrodes in IFG and anterior STG. This was compared to clusters constructed directly from anatomical parcellations. The anatomical clustering produced more variance in all feature dimensions, providing less accuracy, something we will find as well in CHAPTER IV.

This study dives into the spatiotemporal dynamics of ECoG high gamma activity during a speech task, but the feature space across the different dimensions are pre-defined, discretized, and then further reduced to a level that a lot of the fine temporal dynamics are averaged out. We are left with interesting results and insights, but are still

left with findings that are still fairly high level and do not generate insight into how temporal profiles differ across brain regions outside of their temporal activations.

In a different study, (Collard et al., 2016) find clusters, which they term functional network components, that are present during word repetition and picture naming tasks amongst 5 subjects. Interaction strength between pairs of electrodes were estimated with time-varying dynamic Bayesian network models (Song et al., 2009) constructed from high gamma power. Signed principal component analysis (PCA) was used to identify significant electrodes pairings, and thus clusters.

Several task related clusters were observed in the picture naming task. One cluster included interactions within ventral occipital-temporal cortex (VOTC), between VOTC and sensorimotor cortex (SMC) and Broca's area. This cluster became active just after stimulus presentation (50-150ms after stimuli) and peaked in activation 200-600ms after stimuli. A second cluster contained interactions between pSTG or supramarginal gyrus (SMG) and either SMC, Broca's area, or both. This cluster achieved peak activation after the median latency of the subject's spoken response. Similarly, several clusters were found in the auditory word repetition task. One cluster was primarily contained within pSTG and showed activity just after stimulus (150-500ms) and a second smaller significant activity just after median response latency. A second cluster captured interactions between pSTG and Broca's area and peaked after the first cluster, but before subject response. Another cluster was within SMC with activity centered around the median response latency. A final cluster was between SMC and Wernicke's area with

two peaks, pre- and post-response, but with the post-response peak showing more activation.

The auditory word repetition task is similar to the task of our work. Although there is additional preprocessing in this study that removes some of the structure of the neural activity, we will revisit this work in CHAPTER III as we discuss our findings in the context of what was found here. One of the specific method choices that limit the time course information in Collard et al., includes constraining the model to have a lag of 1 sample, or 16ms. This represents the transfer of high gamma information between cortical sites, highly limiting the amount of temporal interaction that can be captured. Additionally, the kernel that was selected for the model was too broad and was non-causal, which may have smeared the directionality of information flow. Lastly, the signed PCAs lose orthomormality and interpretation of explained variance. These design choices were done to allow the analysis to focus on the co-variance between electrodes and maintain interpretability, but have done so at the cost of finer timing details and activity structure.

### *II.3.ii Spectrotemporal Clustering*

The Edward F. Chang lab at the University of California, San Francisco has had a number of studies performing ECoG clustering analysis for speech. Their work amongst these studies has been some of the pioneering work in understanding the higher order aspects of spatiotemporal analysis of speech through ECoG. In particular, three studies are of note (Hullett et al., 2016; Hamilton et al., 2018; Leonard et al., 2019), with each looking at slightly different parts of the speech network. The first study (Hullett et al.,

2016), uses spectrotemporal receptive fields (STRF) and is covered in this subsection, while the other two use cNMF and are covered in §II.3.iii.

Hullett et al. use STRFs to analyze high gamma power from eight subjects listening to continuous speech. The primary finding was that low-level acoustic features of speech propagate from pSTG toward anterior STG (aSTG), corroborating (Chang et al., 2010; Bouchard et al., 2013; Berezutskaya et al., 2017). The STRFs linearly encoded high gamma activity as a weighted sum of stimulus features, both spectral and temporal, over time (Theunissen et al., 2001; Sharpee et al., 2004). STRFs were grouped together using K-means clustering (McQueen, 1967; Lloyd, 1982), with the Silhouette criterion (Rousseeuw, 1987) used to identify the number of clusters. Further preprocessing was done to get modulation transfer functions (MTFs) from the magnitude of the two-dimensional Fourier transform of the STRFs (Singh and Theunissen, 2003). This was done to reduce the empirical complexity, while still maintaining a lot of the underlying structure. An anterior to posterior organization of modulation tuning was found along the STG, with high spectral and low temporal modulation found anteriorly and low spectral and high temporal modulation found posteriorly. This shows pSTG is tuned for temporally fast-changing speech sounds with relatively constant energy across frequency, while aSTG is tuned to temporally slow-changing speech sounds with a high degree of frequency variation.

### *II.3.iii Convex Non-negative Matrix Factorization*

The two other studies by the Chang lab use convex non-negative matrix factorization (cNMF) (Ding et al., 2010) as their primary method for both preprocessing

and clustering. This approach uses the fewest prior assumptions and is therefore the most data-driven approach of prior studies. This allows for maintaining the temporal structure of neural activity and enables insights about the true activity profiles present during speech production.

In the first study discussed here, (Hamilton et al., 2018), the STG again will be the brain region of focus. This study included a larger set of subjects, 27, with a combined total of 2,100 ECoG electrodes. cNMF was used to correlate activity during the task to generate functional brain area clusters. The primary findings of this study are a parallel caudal and rostral partitioning streams that detect stimulus onset and prosodic information, respectively. Of these two dominant activity profiles, one was sensitive to sentence onsets and mainly localized to pSTG (i.e., the caudal stream), while the other had more sustained activity and was localized to anterior and middle STG (i.e., the rostral stream). The silhouette index was used to show the clusters were both functionally and anatomically significant, with the clusters represented 14.5% of the explained variance in the data.

MTFs, similar to (Hullett et al., 2016), were also generated, with similar results supporting the cNMF findings. The caudal cluster had higher temporal and lower spectral modulation selectivity with a preference for long silences, i.e., those found before a sentence, with the inverse in the rostral cluster. The authors postulated that this additional context likely explains the caudal cluster as onset selective, rather than selective for high temporal modulations as found in (Hullett et al., 2016). When looking

at activity latencies, the caudal cluster was also shown to be significantly quicker and shorter than the rostral cluster for onset, peak, offset, and duration.

Electrodes that significantly responded to sentence start were located exclusively in pSTG. In four of the subjects, ECoG electrodes were also placed in the superior temporal plane and lateral surface, including Heschl's gyrus (HG), planum temporale (PT), and planum polare (PP). Activity in the temporal plane were similar to the STG, showing a caudal/rostral distinction, with HG and PT mostly being onset selective and PP having sustained activity. HG differed from the rest of the caudal cluster, however. In addition to sentence onset activity, HG also responded to features throughout the sentence.

The final clustering study, (Leonard et al., 2019), is the closest to our work. This study, again by the Chang lab, built on their prior two studies just discussed. The work by Leonard et al. develops a more detailed view of the spatiotemporal dynamics of speech perception and production.

ECoG high gamma power was computed from 8 subjects performing an acoustic word repetition task. The task progressed through several structured stages: auditory stimulus cue, acoustic stimulus, 2 second pause, another auditory cue, and then subject response. Two unsupervised clustering methods were used to determine functional groupings: 1) cNMF similar to (Hamilton et al., 2018), and 2) k-means. Cluster size was chosen using percent variance explained, with ~90% of the variance determined by 5 clusters. cNMF and k-means produced similar cluster results.

The 5 clusters were separable in their temporal profiles and spatial locations. The 5 clusters are broken up with the following cluster descriptions:

1) This cluster was located in posterior and middle STG, with a small presence in middle temporal gyrus (MTG) and SMG. Activity showed significant responses to listening and production cues and the auditory word (stimulus), but not to the subject's production of target word. The authors comment that this cluster represents short-latency responses to acoustic input.
2) This cluster was located throughout the STG, extending into posterior MTG and SMG and also IFG and dorsal vSMC. Activity displayed significant responses only to speech sounds, both externally and self-generated. The authors comment that this grouping represents short-latency responses to hearing speech (both external and self-produced).
3) The third cluster had a diffuse network across the STG, posterior MTG, inferior posterior parietal cortex, IFG, and vSMC. Significant sustained activity existed during the delay period. The authors comment that this grouping represents phonological working memory.
4) The fourth cluster had similar anatomical coverage as the third cluster, with activity located in STG, posterior MTG, IFG, and vSMC. However, activity was similar to the second cluster. The duration of activity was longer than the second cluster and some activity was present during the delay period. The authors comment that this grouping represents activity evoked by hearing speech and also some contribution to working memory.

5) The final cluster was located primarily within vSMC with a small number of electrodes in posterior STG, MTG, and SMG. Significant activity was consistent with speech production, with activity during the speaking phase, peaking just after onset. The authors comment that this grouping represents motor planning and speech production.

This study is the closest to our work, both in its methods and task. In this study acoustic word stimuli are presented, while ours utilizes visual stimuli. The structure of the task is more prescribed in this study then ours, including auditory cues, which potentially confound some results but provide tighter temporal similarity across trials. Both of our methods aim to limit preprocessing to pull out the structure of the activity. The clustering methods are different, however, with Leonard et al. using cNMF which utilizes a linear similarity distance measure and is a soft clustering technique, while we use a nonlinear similarity measure and enforce hard clustering to generate unique partitions, refer to APPENDIX B for more details on the trade-offs between the approaches. We also go further than this work by quantitatively formalizing a way to describe the activity patterns and compare them. We will return to this work by Leonard et al. in CHAPTER III.

### II.4 Return to Speech

We now return back to a more general discussion of speech to help place ECoG research within the larger field, giving particular attention to several areas: speech modeling, speech timing and propagation, and brain computer interfaces. The goal of this dissertation is to provide additional insights that add new knowledge to our

understanding of the speech network. In doing so, we hope to update our understanding of neural timing and propagation during speech and provide the inputs to enable updating speech networks to get closer to true neural representations. Lastly, we hope our work motivates further research and development in better brain computer interfaces to allow speech to be easier to perform for all.

### *II.4.i Speech Modeling*

The knowledge that has been gained about the speech network through lesion, behavioral, imaging, and electrophysiology studies has been used to craft models of the speech network. Early models relied more heavily on lesion studies, such as the establishment of Broca's area (Broca, 1861, 1865) and Wernicke's area (Wernicke, 1874). Models by Lichtheim (Lichtheim, 1885), (Goodglass, 1993) and Geschwind (Geschwind, 1965, 1979) built on the more simplistic early models to greatly expand the speech network to be more anatomically and functionally widespread. As computers became more readily available, they became useful tools for modeling and simulating the speech network. This allowed for augmented data analyses to go along with data collected in studies to add valuable insight to our understanding of speech, reducing reliance on human subject testing. Some of the notable computer models include (Henke, 1966; Rubin et al., 1981; Dell, 1986; Saltzman and Munhall, 1989; Horwitz et al., 2000; Garagnani and Pulvermüller, 2013).

The Guenther lab at Boston University has an extensive history of studying and modeling speech. A main component of this research has been the development, validation, and maintenance of the DIVA (Directions Into Velocities of Articulators)

neural network model of speech acquisition and production (Guenther, 1994, 1995, 2016; Guenther et al., 1998, 2006; Guenther and Brumberg, 2011; Tourville and Guenther, 2011). This model is a peer reviewed, state-of-the-art representation of neural activity during speech that has been validated to align with findings from human subject tests. For example, the progression and connectivity of activity in premotor, motor, somatosensory, and auditory brain areas of the model have been found to align with prior studies, refer to (Tourville and Guenther, 2011) for a review. Amongst other things, it provides a means test and understand components of speech at a scale and in a controlled manner that cannot be done with human subjects.

Models are an extremely useful tool, as they provide a way to quickly study and perturb a speech network to analyze and gain insights on what is happening without the need, and high cost, of doing so with human subjects. Human subjects are far preferred, but access to subjects for electrocorticography or other invasive studies is strictly limited to those that are undergoing medical procedures who agree to participate in the study. Therefore, models help provide some of the gaps as a useful tool that has unlimited access, but does not fully represent the underlying neural dynamics. This allows for the testing of many different experimental paradigms to see which ones show the greatest potential for new insights before selecting a specific experiment to run on human subjects.

At the same time, caution and care needs to be taken when constructing and updating models based on findings from human subject studies. Utilizing a standard brain representation poses challenges as it is difficult to utilize spatial averaging as

cortical evoked response patterns have been found to be relatively heterogeneous across individuals (Flinker et al., 2011; Leuthardt et al., 2011; Cogan et al., 2014). Thus, we are left with an extremely valuable tool with speech models, as they provide non-invasive ways to invasively explore speech, but the complexity and heterogeneity amongst different individuals creates questions of generality that cannot fully be addressed without more subject data.

### *II.4.ii Timing and Propagation*

Lots of prior work has been conducted to understand the fine timings of the different stages of speech. This prior work provides detailed analyses of the temporal properties of speech. For example, it has yielded understanding of the temporal properties of the STG that includes the onset, offset, peak, and duration of the neural activity (Lerner et al., 2011; Honey et al., 2012; Nourski et al., 2014), how the auditory cortex tracks the envelope of speech (Kubanek et al., 2013), and how higher level semantics, such as sentence structure, are temporally encoded (Halgren et al., 2002; Brennan and Pylkkänen, 2012; Fontolan et al., 2014). Much attention has been given to timing analyses for speech and will not be covered in detail here. Much of this prior work falls within first order analyses, including from ECoG recordings (refer to §II.2.ii). Relevant timing studies will be discussed in the context of the results of this study in their appropriate chapters.

### *II.4.iii Brain Computer Interfaces*

Brain-Computer Interfaces (BCIs) for speech have been improving with better temporal and spatial decoding of neural activity (Guenther et al., 2009; Brumberg et al., 2010), but still lack the performance needed to fully enable an individual to comfortably communicate. ECoG has started to be incorporated into BCIs, but to date has still relied on first order analysis, i.e., §II.2.ii, and thus is temporally bounded in the performance that can be reached. Recent work in ECoG-based BCIs include using high gamma power for auditory attention (Brunner et al., 2017) and phoneme recognition and prediction (Moses et al., 2016). We hope that moving to higher order analyses, such as in our work amongst others, will help provide some inspiration and lead to some of the needed gains of BCIs to improve their accuracy and enable them to be life changing technologies for many individuals.

## II.5 Situating this Research

Much of our current understanding of speech comes from lesion or behavioral studies, which lack insight into the neural underpinnings driving the observations, or from imaging studies that rely on methods with time courses slower than that of speech. For example, fMRI has temporal measurement accuracies around a second, while speech time courses are closer to 100 ms. This results in the fine temporal details and the propagation of neural activations across brain areas to get smoothed out, losing some of the important temporal details. This is an area that ECoG has helped to fill. However, much of the current ECoG work has been limited to lower order understandings, such as

finding electrodes that have activity statistically different from a baseline period and only performing simple characterizations, e.g., onset, peak, and duration. This type of analysis loses the detail of activity between these discrete points, such as how activity builds from the onset to the peak of activity. This additional information that is not currently being characterized has the potential to provide insights into the underpinnings of speech and provide a means to better understand *how* different brain regions respond similarly or differently to a task.

The last few years have seen new methods starting to emerge that do not remove this important temporal information. These methods have been primarily centered around clustering analysis studies, with (Leonard et al., 2019) leading the way with the use of cNMF. In this work, we aim to further the developments within this emerging research area, while building on the lessons learned from the studies highlighted in §II.2 and §II.3. We take a unique approach to clustering that is designed with the properties of ECoG in mind. Specifically, the clustering method utilizes a nonlinear distance measure to emphasize similarity between important activity time points, i.e., activity peaks, and performs hard clustering to create unique group assignments. Further we are the first to formalize a method to quantitatively explain activity patterns, providing a powerful framework to not only describe our findings, but also to open up new avenues of research.

# CHAPTER III: Canonical Speech Clusters

## III.1 Significance

The identification of characteristic time courses of neuronal activity during movement planning and execution has provided critical insights into the brain mechanisms underlying motor control in non-human primates. Due to the relative lack of electrical recordings from the human brain, little is known about the temporal profiles of neuronal populations involved in uniquely human acts such as speech. In this study we identify and quantitatively characterize eight canonical time courses of neural activity recorded using electrocorticography during production of simple speech utterances. The resulting time courses provide unprecedented detail regarding the nature and timing of neural computations underlying the translation of phonological information into motor and acoustic output.

## III.2 Introduction

Intracranial electrophysiology has resulted in a wealth of information concerning the time course of neural computations underlying the control of movements in non-human primates. The activity patterns of neurons involved in arm movement control have been shown to fall into classes that exhibit characteristic temporal profiles, including phasic responses that quickly rise and then return to baseline immediately preceding movement onset or coincident with movement execution, tonic responses that change relatively abruptly from one activity level to another near the time of movement onset, and ramp responses that gradually increase over the course of a movement (Cheney and

Fetz, 1980; Kalaska et al., 1989). These characteristic time courses differentiate functional classes of neurons involved in different stages of movement planning and execution, and they illuminate the neural computations performed during these stages. The scarcity of high-resolution intracranial recordings in humans has thus far precluded such a description of brain activity during human movement execution. The vast majority of studies of neural activity during human movements have involved non-invasive technologies such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) that lack the temporal resolution necessary to characterize the time course of electrical activity underlying the various stages of movement planning and execution; the temporal resolution of blood flow responses measured by PET and fMRI are on the order of seconds, while the production of a phoneme is on the order of 100 ms. Against this background, the primary purpose of the current study was to identify and quantitatively describe characteristic time courses of electrical activity in the brain during the production of simple speech movements using recordings from surface and depth electrocorticography (ECoG) electrodes implanted for pre-surgical mapping in individuals with intractable epilepsy. Although ECoG does not capture activity at a single neuron level like primate electrophysiology, it affords the highest combination of spatial and temporal resolution of any relatively widely used human neural recording technology. Furthermore, high gamma power (broadly interpreted as some portion of the 60-500 Hz range) in the ECoG signal correlates with local neural activity (Ray et al., 2008; Miller et al., 2009) as well as motor behavior (Bouchard et al., 2013) and auditory

processing (Kubanek et al., 2013; Hullett et al., 2016; Berezutskaya et al., 2017), making it an excellent source for studying local time courses of speech-related neural activity.

A number of prior ECoG studies have investigated neural activity (usually in the form of high gamma power) during speech motor control, revealing many novel insights. For example, several studies (Bouchard et al., 2013; Lotte et al., 2015; Conant et al., 2018) described time courses and spatial distributions of articulator-specific activations during speech. (Martin et al., 2014; Angrick et al., 2019; Anumanchipalli et al., 2019) applied signal decoding techniques to reconstruct acoustic signals recorded during spoken utterances from ECoG recordings. (Mugler et al., 2018) identified distinct articulatory and phonemic representations in the motor cortex and inferior frontal gyrus, respectively. (Brumberg et al., 2016) examined cortical activity during sentence repetition to identify a global trend in which frontal-motor activations precede auditory cortical activations, with the former beginning approximately 440 ms prior to vocal onset and the latter extending 440 ms beyond vocal onset (see also (Leuthardt et al., 2012)). In one of the first attempts to identify characteristic time courses during speech that are common across subjects, (Leonard et al., 2019) used a soft clustering procedure on ECoG data from the left hemisphere collected during a cued word repetition task involving auditory stimuli and identified five characteristic clusters: one that responded only to the auditory stimulus, two that responded to both the auditory stimulus and the production period, and one that responded only during production.

The current study extends this prior work in several ways. First, we employ a novel Kalman filter-based trend analysis with an unsupervised clustering approach to

identify change points between high gamma activity trends (for example, between the flat baseline trend and the onset of task-related activity, or between an activity plateau and the return to baseline) and to quantitatively characterize these trends, thereby providing a significantly more detailed account of the characteristic shapes of activity time courses. Second, we utilize recordings from both cerebral hemispheres, allowing us to investigate hemispheric differences in word production (most evident in premotor and prefrontal cortical regions). Third, we utilize orthographically presented monosyllabic stimuli rather than auditory stimuli or sentence stimuli, thereby extending the study of characteristic time courses during speech to the commonly used experimental paradigm of single word reading.

## III.3 Materials and Methods

### *III.3.i Participants*

Data were obtained from five neurosurgical patients (4 males, 1 female) undergoing surgical treatment of medically intractable epilepsy; refer to Table 1 for more details. Written informed consent was obtained from all subjects, and all research protocols were approved by the appropriate institutional review board.

**Table 1: Subject Information**

| Subject | Handed | Age (yrs) | Hemi-sphere | Num. Trials | Response Latency (ms) | Seizure Focus | Sex |
|---|---|---|---|---|---|---|---|
| **S357** | 90+ | 37 | L | 108 | 1004.2 +/-209.3 | L. mesial temporal | Male |
| **S362** | 100+ | 60 | L & R | 281 | 894.7 +/-202.4 | R. Ant Middle Frontal Gyrus, L Post Inf Parietal, l Frontal Pole | Male |
| **S369** | 100+ | 30 | R | 118 | 830.3 +/-158.7 | R mesial temporal [para-hippocampal gyrus, amygdala, hippocampus, fusiform gyrus] | Male |
| **S372** | 75+ | 34 | L | 144 | 860.0 +/-121.2 | L temporal pole, Parahippocampal gyrus | Male |
| **S376** | 90+ | 49 | R | 133 | 1027.2 +/-261.1 | R Parahippocampal gyrus | Female |

### *III.3.ii Experimental Design*

Subjects read aloud orthographic stimuli projected onto a video screen. The stimulus set used in this study consisted of consonant-vowel-consonant (CVC) pseudowords constructed from the combinations of four consonants (/b/, /d/, /g/, and /ʤ/) and 3 vowels (/æ/, /i/, and /u/) to generate 12 CVCs: /bæg/, /big/, /bug/, /dæʤ/, /diʤ/, /duʤ/, /gæb/, /gib/, /gub/, /ʤæd/, /ʤib/, and /ʤud/. A brief practice session was used to familiarize subjects to the orthographic representation of each stimulus. Each collection period (run) consisted of 72 CVCs grouped into 36 pairs. For each pair, the first CVC was presented on the screen for 1 second, followed by a gap of 1.5 seconds before the second CVC was presented for 1 second. The time between word pairs was randomly drawn between 3, 4, or 5 seconds. Subjects were instructed to say each stimulus as soon as it was presented, i.e., there was no "go signal" between the reading portion and the

speaking component. The analyses in the current study utilized data from only the first word in each pair to minimize potential residual effects from prior productions on the ECoG signal. After an introductory period to familiarize the subject with the experimental protocol, each subject participated in 3 or 4 36-pair runs.

### *III.3.iii Instrumentation*

A condenser microphone (Beta 87C, Shure, Niles, IL) captured each subjects' speech, which was amplified (MK3, Mark-of-the-Unicorn, Cambridge, MA) and passed into a multi-channel data acquisition system (DAS; System3, Tucker Davis Technologies, Alachua, FL, or Atlas, Neuralynx, Bozeman, MT) that also simultaneously collected TTL signals denoting presented visual stimuli and ECoG signals (see §III.3.iv). We utilized an online sampling rate of >12kHz for voice signals but resampled to 12kHz offline in MATLAB (MathWorks, Natick, MA).

### *III.3.iv Electrocorticography Acquisition*

Research recordings were initiated after the subjects had fully recovered from electrode implantation surgery. Subjects were awake and sitting comfortably in bed during all experimental recordings. Subdural implantation of the electrode arrays allowed for ECoG signals to be directly recorded from the cortical surface. The ECoG signals were filtered (1.6–1000 Hz anti-aliasing filter), digitized with a sampling frequency of >2000 Hz and then resampled offline in MATLAB.

### *III.3.v Electrode Implantation and Localization*

The devices used to record electrical activity of the brain were a combination of surface (i.e., subdural) and penetrating depth multi-contact electrode arrays[1]. Each surface array consisted of platinum-iridium disc electrodes arranged within a silicone sheet (Ad-Tech, Racine, WI or PMT, Chanhassen, MN). The distance from the center of one electrode to the center of an adjacent electrode measured 5 or 10 mm, while each individual electrode had a contact diameter of 3 mm. Depth electrodes were utilized in all subjects with placement locations dictated by clinical needs of each subject. The extent of the array coverage varied between subjects due to the different clinical considerations specific to each subject. After surgical implantation, subjects were continuously monitored via video-EEG during a fourteen day hospitalization to correlate seizure activity with brain activity for purposes of epilepsy treatment. During this period, high resolution monitoring verified that cortical areas relevant to this study did not show abnormal inter-ictal activity. Once this two week monitoring period was complete, the electrodes were surgically removed and the localized seizure focus was resected.

High-resolution digital photographs were taken intra-operatively during electrode placement and removal. In addition, pre- and post-implantation MR (0.78×0.78×1.0 mm voxel size) and CT (0.45×0.45×1.0 mm voxel size) scans were conducted. This information was combined to localize the exact position of the recording electrodes in

[1] Results presented come from both surface and depth electrodes, with roughly a third of the electrodes being from depth arrays. The analysis was rerun only using the surface electrodes, which is similar to prior studies, to confirm that the same results are obtained. These results are presented in APPENDIX D.

each subject. FMRIB's Linear Image Registration Tool was utilized to apply a three dimensional rigid fusion algorithm that successfully allowed pre- and post- implantation CT and MRIs to be co-registered (Jenkinson et al., 2002). The coordinates for each electrode from post-implantation MRI volumes were transferred to pre-implantation MRI volumes, allowing the relative location of each individual electrode contact in relation to surrounding distinguishable brain structures to be compared in both the pre- and post-implantation MRI volumes. This comparison is helpful for improving the accuracy of electrode localization since implantation causes medial displacement of the cerebral hemisphere, which leads to greater deviation of the superficial cortex compared to deeper structures. The resultant electrode positioning was then mapped onto a three dimensional surface rendering of the lateral surface that was specific to the architecture of each subject's brain. The estimated spatial error rate when localizing these electrodes is less than 2 mm.

Electrode locations are provided in Figure 1, with all electrodes across all subjects plotted on the FreeSurfer (Fischl, 2012) common reference brain (panel b) and individual subject electrode locations plotted on the subject's own magnetic resonance imaging (MRI) scan (panel c). A total of 1036 electrodes were analyzed across the 5 subjects.

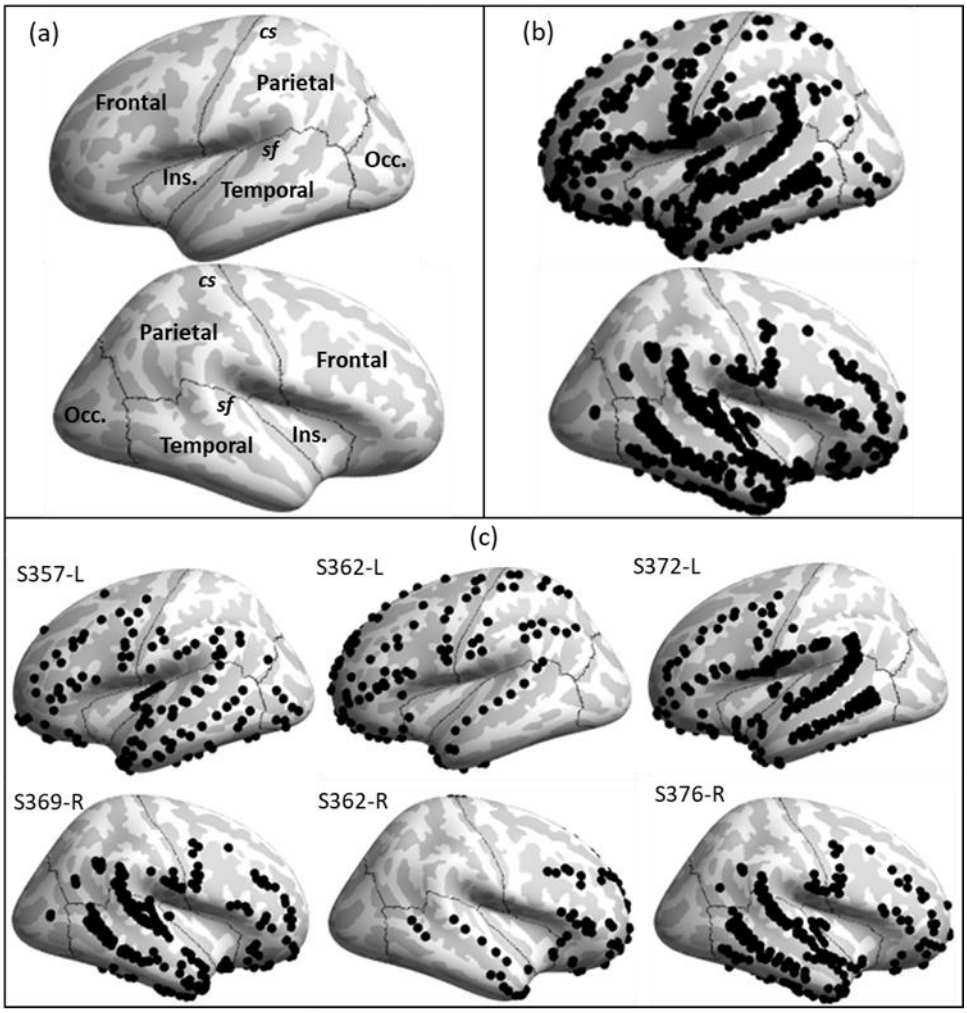


**Figure 1: Electrocorticographic Recording Locations on Inflated Cortical Surfaces**

(a) Reference brain template with lobular labels. Central sulcus (cs) and Sylvian fissure (sf) denoted. (b) Electrodes from all subjects plotted together on common brain. (c) Individual subject electrode locations.

### *III.3.vi Audio Preprocessing*

Speech onset was measured using a semi-automated method. A 20 ms rectangular kernel was convolved with the absolute value of the recorded audio signal. Resulting values that were above an empirically determined threshold were marked as periods of voicing. Coarse onset estimates were determined to be at the beginning of any contiguous period that exceeded the threshold and with a gap greater than 300 ms from

the previous onset. Manual verification and correction was performed using the Praat software suite (www.fon.hum.uva.nl/praat/) to determine onset validity and refine the location of voicing onset. Average audio signals were computed to get sound envelopes by taking the root mean squared (RMS) value of the raw audio over 50 ms time windows (Kubanek et al., 2013).

### *III.3.vii Neural Signal Processing*

The recorded data were downsampled to 1 kHz for further processing with a polyphase anti-aliasing filter using the *resample* function in MATLAB. After downsampling, the DC component was independently removed for each channel (electrode) by subtracting the average value for the channel over the entire collection period. Line noise was removed using notch filters at 60 Hz and harmonics. This was done using the *tmullen_cleanline* function in EEGLAB (Bigdely-Shamlo et al., 2015), which builds on the Chronux toolbox.

Next, bad channels were identified and removed from further analyses. Bad channels consisted of two types: a) those that were clinically or experimentally determined to be invalid, and b) those that were labeled invalid during preprocessing. For the latter, a kurtosis analysis was performed to remove channels that were corrupted by noise or were unexpectedly peaky, such as with eye blink artifacts (Mognon et al., 2011; Tuyisenge et al., 2018). Channels identified for removal were manually verified. Signals were then re-referenced according to a Common Average Reference (CAR) scheme (Crone et al., 2001a), with electrodes averaged across each grid of electrodes to remove non-neural noise artifacts from the shared collection hardware.

Following numerous prior ECoG studies (e.g., (Edwards et al., 2009; Chang et al., 2011)), we focused our analyses on the time course of signal power in the high gamma frequency range[2]. Specifically, the log-analytic amplitude of the Hilbert transform was used to bandpass filter the ECoG recordings into 8 logarithmically spaced bands spanning 70 – 150 Hz (cf. (Moses et al., 2016)). The analytic signal was computed for each band and the absolute value was taken as the analytic amplitude, which represents the envelope of the bandpass filtered signal. These amplitudes were then averaged together to get a log-analytic amplitude representation.

After filtering, trials were extracted and re-referenced to a baseline. Trial extraction was performed under two alignment conditions: 1) visual presentation of the stimulus and 2) voice onset of the speech response. These alignment segmentations are referred to as stimulus presentation and voice onset, respectively. The stimulus presentation epochs had a trial duration of 3 seconds, starting 1 second prior to stimulus presentation and lasting 2 seconds after (average voice onset started 916 ms after stimulus presentation, with a 208 ms standard deviation). The voice onset epoch also had a 3 second trial duration, starting 2 seconds prior to voice onset and ending 1 second after (with average stimulus presentation 916 ms prior to onset). A baseline period before stimulus presentation was used to re-reference the signal. The baseline period was taken to be the first 500 ms prior to stimulus presentation, and the high gamma signal at each time point in the trial was re-referenced as a z-score relative to the trial's baseline period

[2] Additional analysis was performed looking at the beta frequency band. The results of this analysis can be found in APPENDIX D.

(Edwards et al., 2010). All trials within an alignment condition were averaged to create an event-related spectral perturbation (ERSP) (Makeig, 1993) that captures the average response of high gamma power for the alignment condition.

Next, electrodes that had a significant ERSP response were identified and kept for further analysis. First, electrodes that did not deviate beyond a 95% confidence interval of baseline activity were marked as non-significant and removed from further analysis. Remaining electrodes were then subjected to a Kalman filter-based trend analysis (described further in §III.3.ix below); only electrodes that deviated from the data-driven baseline trend identified by the Kalman filter were kept for further analysis. The 1036 total electrodes from the five subjects were reduced to 319 significant electrodes for the stimulus presentation alignment case and 334 for the voice onset case.

More details about the general methods, including motivation and alternatives considered can be found in APPENDIX C.

### *III.3.viii Statistical Analysis*

Clustering was performed to generate electrode groupings using pairwise distances measured for each alignment case. Pairwise comparisons of ERSP responses (after normalizing each ESRP response to range from 0 to 1) were computed using a distance measure that emphasized activity differences further away from the non-task baseline. Specifically, an exponential difference between signal values was computed using Equation 1:

**Equation 1: Pair-wise Electrode Exponential Distance Function**

$$DIST = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(e^{p_i} - e^{q_i})^2}$$

where $p_i$ and $q_i$ are the signals on electrodes $p$ and $q$ at time point $i$ out of $N$ total samples. For simplicity, we drop the 1/N common normalization term. This distance measure emphasizes significant activity time points and hence puts more weight on the similarities or differences of these time points. This differs from most prior studies, which characterized electrode signal similarity using linear measures (correlations) that put equal weight on non-significant time points rather than focusing on similarities or differences during key time points of the activity such as peaks or plateaus.

A hybrid clustering method that combines partitioning and hierarchical clustering was used to identify electrodes that displayed similar time courses according to the distance measure just described (Warren Liao, 2005; Aghabozorgi et al., 2015). This approach initially assigns each electrode to its own cluster, as in hierarchical agglomerative clustering. At each iteration step the pairwise distance between each cluster is computed. The two clusters with the closest match are then merged. Merging consists of re-computing an average for all electrodes that are members of the cluster, which results in a new cluster centroid. This hierarchical approach by itself creates a non-monotonic cluster tree. To ensure a monotonic cluster tree, a partitioning refinement step is taken to look at any cluster that has a closer distance measure in the new cluster representations compared to the distance measure of the clusters just merged. The partitioning step reallocates electrodes between the two merged clusters and any clusters

breaking the monotonic relationship, generating clusters that maintain a monotonic cluster tree. This process repeats for each step of the iterative clustering until all electrodes are merged into a single cluster. This clustering method combines the strength of cluster tree generation through hierarchical clustering methods with the ability to maintain a monotonic cluster tree to enable the selection of the number of clusters. The partitioning refinement step functions similar to k-means over a subset of the electrodes.

After this clustering procedure, the number of clusters that best capture the true nature of the underlying data was selected. A distance threshold can be set to select the number of clusters from the cluster tree. Since there is no clear method for choosing a threshold, two different methods were employed to choose the most informative number of clusters. First, the "elbow" method (Thorndike, 1953) was used to select the number of clusters based on the elbow in the cluster tree, which looks at the distances between clusters at each branch of the tree. More precisely, we selected the elbow from the derivative of this function, which was far more pronounced. This elbow, where the derivate shows a very noticeable slowing rate of change in the reduction in distance that additional clusters would add, occurred at six clusters for both alignment cases. In the second method, the percent variance explained using a comparison of the sum of squares of within-cluster variance to total variance was calculated (Goutte et al., 1999). This method also indicated that six clusters in each alignment case provided the best account of the data. This choice of clusters explains 73% of the variance for the stimulus

presentation case and 69% for the speech onset case, with additional clusters only marginally adding to the explained variance[3].

In addition to clustering the data separately for the stimulus onset and voicing onset alignment cases, the clustering procedure was also employed to identify clusters that appeared in both alignment cases. The average offset between stimulus presentation and speech onset (916 ms, with a standard deviation of 208 ms), was used to align the two cases. Using this combined dataset, the 12 clusters found from the two alignment cases (6 from each) were found to reduce to an aggregate set of 8 clusters. Four of these clusters were seen in both alignment cases and four were only seen in one of the alignment cases. The results of this analysis were used only to identify clusters that appeared in both alignment cases; the clusters described in the results section, §III.4, are from the individual alignment cases rather than from this combined clustering analysis.

More information about the clustering analysis, including motivation and alternatives considered can be found in APPENDIX B.

#### *III.3.ix Trend Analysis*

A novel data-driven statistical method was used to identify trends and change points in high gamma traces for two purposes: (1) to identify electrodes in which activity changed significantly from the baseline (refer to §III.3.iv and §III.3.vii), and (2) to quantitatively describe the shapes of the characteristic time courses resulting from the cluster analysis. Past studies have utilized functional representations to capture changes

[3] Only clusters that existed across multiple subjects are discussed in the results. All results presented had electrodes present in at least 3 of the 5 subjects. Results from the clusters that constrained to only one subject are presented in APPENDIX D.

in neural temporal dynamics, such as splines (e.g., (Brumberg et al., 2015)) which provide piecewise linear breakdowns for trend analysis. We instead utilized a dynamic method that detects trend changes in the data with fewer priors on the form the changes can take. The method is based on detecting change points (Page, 1963; Aminikhanghahi and Cook, 2017) with a Kalman filter (Kalman and Bucy, 1961). A Kalman filter is a statistical method that estimates the internal state of a linear dynamic system from a series of measurements that include process noise (in our case, error inherent to the neural signal model) and observation noise (noise inherent to the ECoG recording process).

For our trend analyses, the Kalman filter estimates high gamma power ($g$) and its time derivative (or *trend*, $\dot{g}$) to model the 2-dimensional state vector $\boldsymbol{X} = [g, \dot{g}]^T$, where $^T$ is the transpose function. The state transition matrix, $\boldsymbol{A}$, captures the relationship between these states[4] at each time point $i$. The Kalman prediction, $\widehat{\boldsymbol{X}}(:, i)$, is based on this state transition matrix and the filter's value at the prior time point. The model of the system, $\boldsymbol{X}(:, i+1)$, is similarly modeled, with the inclusion of the realized noise term, $w$. These equations are collected in Equation 2:

**Equation 2: State Transition Matrix, System Model, and Kalman State Prediction**

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\boldsymbol{X}(:, i+1) = \boldsymbol{A}\,\boldsymbol{X}(:, i) + w$$

$$\widehat{\boldsymbol{X}}(:, i) = \boldsymbol{A}\,\widehat{\boldsymbol{X}}\boldsymbol{X}(:, i-1)$$

[4] Here we use a simple linear estimation procedure. More complex filters were tested, but the linear state transition matrix performed best and was the most parsimonious.

The covariance (or uncertainty) of the estimate, termed $\boldsymbol{P}$, is calculated using Equation 3:

**Equation 3: Prediction Covariance Estimate**

$$\boldsymbol{P} = \boldsymbol{A}\,\boldsymbol{P}\,\boldsymbol{A}^T + \boldsymbol{Q}$$

where $\boldsymbol{Q}$ is the covariance of the process noise, i.e., the noise present in the underlying neural activity. For the first time step of the baseline period, $\widehat{\boldsymbol{X}}$ is initialized to $[0\ 0]^T$, $\boldsymbol{Q}$ is initialized to difference-stationary whitened variance during the baseline period, and $\boldsymbol{P}$ is initialized to $\boldsymbol{Q}$. Together the calculations above are called the *prediction step*.

After predicting the state of the system based on its prior estimated state in the prediction step, $\widehat{\boldsymbol{X}}(:, i)$, the Kalman filter then updates this estimate based on the observation $Z(i)$, which is the high gamma power measured by the electrode at the current time point. The rate of change of the measured power is also estimated as the difference between the power at the current time point and the power at the last initialization point divided by $\Delta$, which is the number of time steps since the last initialization point. The *update step* is governed by Equation 4:

**Equation 4: Kalman Filter Update Step**

$$\widehat{\boldsymbol{X}}(:, i) = \widehat{\boldsymbol{X}}(:, i) + K\left(Z(i) - \boldsymbol{H}\,\widehat{\boldsymbol{X}}(:, i)\right)$$

where $K$ is the *Kalman gain* that determines the relative weight to put in new observations versus the prediction, and $\boldsymbol{H}$ is the measurement model that maps the model state space $\widehat{\boldsymbol{X}}$ into the observation space $Z$. In our case $\boldsymbol{H}$ is set to $[1\ 0]^T$ since only the power is observed. The Kalman gain is calculated as follows:

**Equation 5: Kalman Gain Term**

$$K = \alpha \, \boldsymbol{P} \, \boldsymbol{H}^T (\boldsymbol{H} \, \boldsymbol{P} \, \boldsymbol{H}^T + R)^{-1}$$

where $R$ is the covariance of the observation noise, which is initialized to be the overall variance in the power of the baseline period, and $\alpha$ is a decay factor that gives less weight to new observations as time persists and evidence is gained for a given trend according to Equation 6:

**Equation 6: Learning Rate as Decay Factor**

$$\alpha = e^{-\Delta/(\delta F_S)}$$

where $F_S$ is the sampling rate and $\delta$ is a time constant set to 100 ms. The parameter $\alpha$ functions to "freeze" trends as evidence for the trend accumulates, which in turn allows deviations from the trend (change points) to be identified before the model is corrupted by data that does not fit the trend.

To identify change points, a threshold is set for how far away a new observation, $Z(i)$, can be from its estimate, $\widehat{\boldsymbol{X}}(1, i)$. The threshold is set based on the empirical variance across the electrode's z-scored baseline period and only takes into account the power term. An inverse Q-function is used to get the 95% confidence value for the variance in the baseline power for each electrode. This results in a beta distribution across all electrode confidence values, which are all z-scored to have the same statistical representation. A more stringent 99% confidence value is used to select the threshold to use from this distribution, resulting in a threshold that is representative across all electrodes and not electrode-specific, and thus correcting for multiple comparisons. If a

significant change in trend is detected at any point after the baseline period, the electrode is deemed to be responsive to the task and is included in the cluster analysis.

Each time a change point is detected, a new Kalman filter is initialized similar to the original one started on the baseline period, but now the data used to initialize the filter is from the time of the change. Values are re-initialized from priors or what was found in the baseline period, as discussed above, with two exceptions. First, the estimate covariance, ***P,*** is recalculated using the current values at the time point of the signal, as prior studies have found that there are changing dynamics during an ECoG task that are dependent on the activity being captured, such as a reduction in variability during stimulus onset (Dichter et al., 2016) and an increased variance with increasing response amplitudes (Tolhurst et al., 1983; Ma et al., 2006). Second, the empirical trend is recalculated using 100 ms of data around the change, with 10% of points in the past and 90% in the future of the change. The decay rate is reset to allow the Kalman filter time to re-learn the new trend before it becomes "frozen".

Using this approach, two trend patterns were identified in the characteristic clusters, referred to as *symmetric* and *ramp* shapes. These characterizations come from two specific trends within a high gamma power trace. The first is the *onset* trend, which is the initial upward trend starting from the first change point after the baseline period. The second is the *offset* trend, which captures the corresponding decrease in high gamma power to return back to baseline values. This return to baseline occurred either immediately after the onset trend or following a period of sustained activity characterized by a separate trend that captured a plateau in high gamma activity. Traces in which the

rate of increase in the onset trend was within 10% of the rate of decrease of the offset trend were characterized as *symmetric*. If the onset and offset rates varied by greater than 10%, they were considered asymmetric. Assessment of the asymmetric traces revealed that, in all cases, the onset trend was significantly faster than the offset trend, a pattern characterized herein as a *ramp* activity pattern.

More details on the Kalman filter change point detection used for trend analysis, including motivation and alternatives considered, can be found in APPENDIX A.

## III.4 Results

Participants read aloud CVC words as they appeared on a video display. Two time points were used to align high gamma power traces across trials: (1) onset of the visual orthographic stimulus, and (2) onset of the vocal response. A hybrid clustering algorithm was then used to identify characteristic time courses (clusters) that were common to at least 3 of the 5 participants. For each alignment point, high gamma traces fell into six characteristic time courses (clusters), with additional clusters adding only a small amount of explained variance. The resulting clusters are shown in Figure 2 for (a) stimulus and (b) vocal onset alignments. Summaries of activation onset times and peak activation times for these clusters are provided in Table 2. An aggregate clustering step across the 2 alignment cases was run to identify clusters that were common to the two time alignment points, reducing the 12 clusters from the two alignments to a set of 8 *canonical time courses*, 4 of which occur in both time alignment cases. These canonical time courses fell into 4 groupings that approximately align with speech production processing stages,

although this knowledge of the speech network was not built into the analysis methods. The following subsections describe the identified clusters in further detail, organized by speech processing stage: (i) early stimulus processing, (ii) phonological-to-motor processing, (iii) motor execution, and (iv) auditory processing.

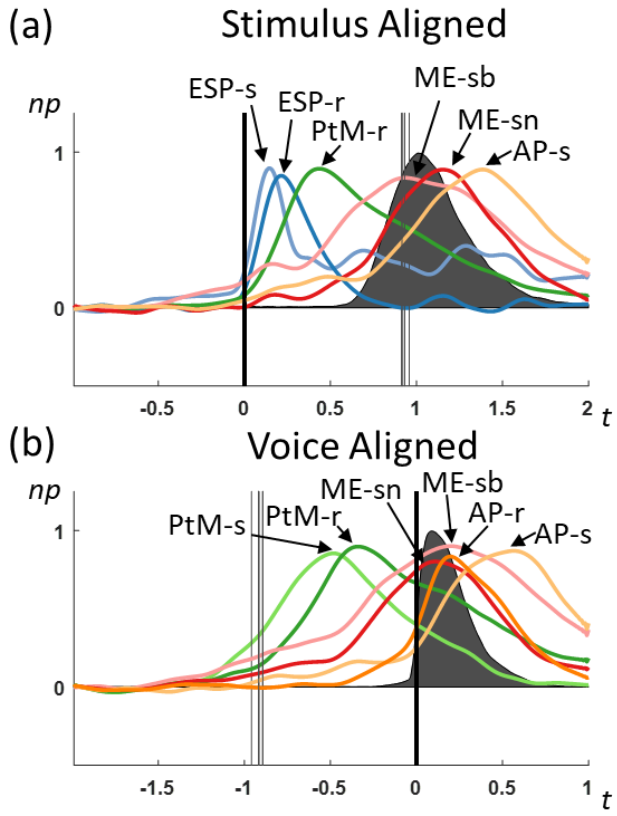


**Figure 2: Characteristic Time Courses During Speech**

High gamma power (normalized power (np)) time courses (time (t) in seconds) across two alignment conditions. Average audio signal amplitude indicated by gray shaded region. (a) Six characteristic time courses from stimulus presentation aligned case: Early Stimulus Processing – symmetric (ESP-s), Early Stimulus Processing – ramp (ESP-r), Phonological-to-Motor – ramp (PtM-r), Motor Execution – symmetric broad (ME-sb), Motor Execution – symmetric narrow (ME-sn), and Auditory Processing – symmetric (AP-s). Stimulus presentation occurred at t=0, shown with vertical solid black line. Average voicing onset per cluster shown in fainter vertical black lines (different lines since not all subjects showed a response for all clusters). (b) Six characteristic time courses from voicing onset aligned cases: Phonological-to-Motor – symmetric (PtM-s), Phonological-to-Motor – ramp (PtM-r), Motor Execution – symmetric narrow (ME-sn), Motor Execution – symmetric broad (ME-sb), Auditory Processing – ramp (AP-r), and Auditory Processing – symmetric (AP-s). Time axis reference to voicing onset (t=0) with a solid vertical line for the time of voicing onset. Fainter vertical lines for average stimulus presentation time per cluster.

**Table 2: Cluster timing landmarks (start, peak, and end times, in ms) and onset/offset ramp slopes (in units of normalized power rate) for the 8 canonical activity patterns by alignment case.**

| | Stimulus Alignment | | | | | Voicing Alignment | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Cluster | Start (ms) | Onset (1/s) | Peak (ms) | Offset (1/s) | End (ms) | Start (ms) | Onset (1/s) | Peak (ms) | Offset (1/s) | End (ms) |
| Early stimulus processing - symmetric | 10 | 4.30 | 150 | -3.98 | 310 | | | | | |
| Early stimulus processing - ramp | 40 | 3.66 | 210 | -2.03 | 530 | | | | | |
| Phonological-to-motor processing - symmetric | | | | | | -1040 | 1.16 | -480 | -1.08 | 320 |
| Phonological-to-motor processing - ramp | 110 | 2.21 | 440 | -0.79 | 1410 | -830 | 1.27 | -340 | -0.70 | 800 |
| Motor execution - symmetric broad | 80 | 0.78 | 930 | -0.86 | 1940 | -710 | 0.72 | 200 | -0.69 | >1000 |
| Motor execution - symmetric narrow | 640 | 1.35 | 1150 | -1.31 | 1690 | -380 | 1.13 | 110 | -1.20 | 640 |
| Auditory processing - ramp | | | | | | -50 | 2.07 | 200 | -1.10 | 680 |
| Auditory processing - symmetric | 760 | 1.24 | 1380 | -1.14 | >2000 | 10 | 1.35 | 570 | -1.33 | >1000 |

### *III.4.i Early Stimulus Processing*

The first grouping consisted of two canonical time courses that showed brief activity immediately following the onset of the visual orthographic stimulus, as shown in Figure 3. Both time courses were found only in the stimulus-aligned analysis. In the first cluster (Figure 3a), activity starts 10 ms after stimulus onset and peaks at 150 ms, falling back to near baseline by 310 ms. The activity onset and offset rates (as characterized by the Kalman filter trend analysis) showed a symmetric pattern; this canonical response is thus referred to as *Early Stimulus Processing – Symmetric (ESP-s).* Interestingly, activity remains slightly above baseline after the offset trend for the duration of the trial. Electrodes in this cluster were found near the anterior junction of the temporal and frontal lobes in the left hemisphere and in right posterior middle and inferior temporal cortex (see Figure 3).

In the second cluster (Figure 3b), activity rapidly increases approximately 40 ms after stimulus presentation and decays more slowly back to the baseline before speech vocalization starts. Activity is slightly broader in duration then the symmetric cluster and

peaks later, at 210 ms. Because high gamma power shows an asymmetric activity pattern with a faster activity increase than activity decrease, this canonical time course is termed *Early Stimulus Processing – Ramp (ESP-r).* Electrodes in the ESP-r cluster were found in posterior middle and inferior temporal cortex bilaterally and in left frontal cortex near the junction of the precentral sulcus and the inferior frontal sulcus.

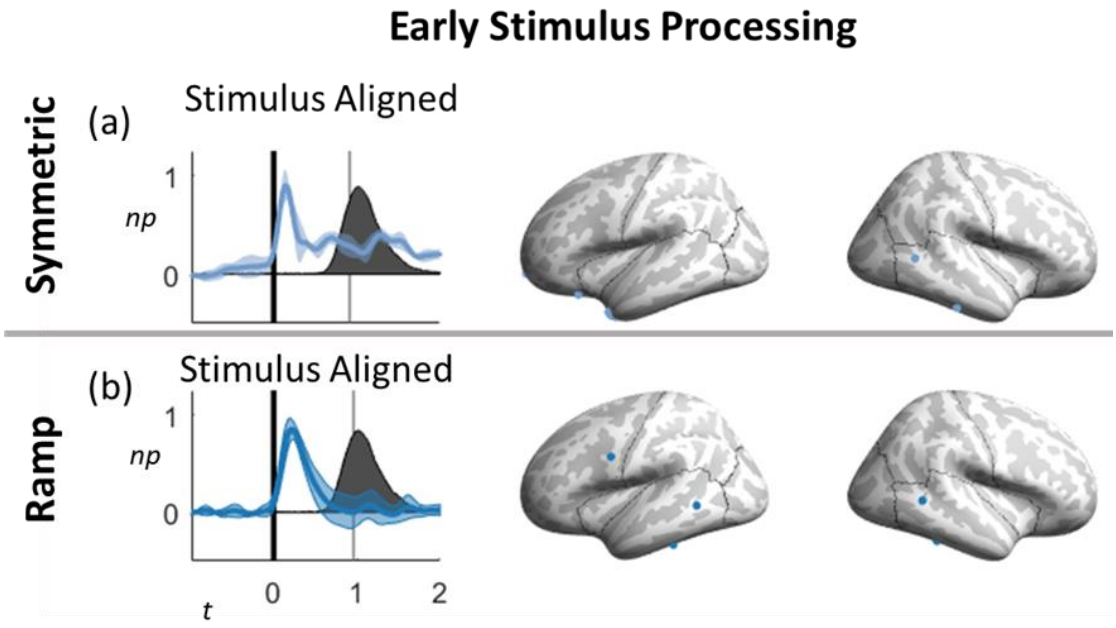


**Figure 3: Early Stimulus Processing Clusters**

Two characteristic time courses included: (a) symmetric and (b) ramp. High gamma cluster activity patterns are shown on the left (normalized power (np) over time (t), in seconds), electrode locations for clusters on the right. Solid vertical lines indicate alignment condition in which the cluster was identified, and fainter vertical lines show average location of other alignment condition. Average audio signal amplitude indicated by gray shaded region.

### *III.4.ii Phonological-to-Motor Processing*

Three clusters, falling into two canonical time courses, had activity peaks that were located approximately halfway between stimulus onset and voicing onset. Based on this timing, combined with the observation that the large majority of the electrodes exhibiting these time courses were in the left frontal cortex (heavily associated with

language production and speech planning processes; e.g., (Broca, 1861, 1865; Guenther, 2016; Friederici, 2017)) and posterior inferior temporal cortex (a site associated with word and nonword reading; (Vigneau et al., 2005)), these canonical time courses are called *Phonological-to-Motor Processing – Symmetric (PtM-s)* and *Phonological-to-Motor Processing – Ramp (PtM-r).* The PtM-s time course (Figure 4a) was found only in the voice-aligned analysis. In this cluster a gradual activity ramp-up starts roughly at stimulus presentation (1040 ms prior to voice onset) and reaches a peak 480 ms before voice onset, then gradually decays back to baseline approximately when production is ending. The activity pattern has a symmetric shape, though it is far more temporally broad than the ESP-s cluster.

The PtM-r time course was exhibited by a cluster in each of the two alignment cases (Figure 4b,c). This time course starts with a relatively rapid activity increase just after stimulus onset and peaks 440 ms after stimulus presentation in the stimulus-aligned case (Figure 4b) and 340 ms before voice onset in the voice aligned case (Figure 4c), followed by a gradual return to baseline at 800 ms after voice onset, when production is complete.

In addition to left frontal cortex (including the supplementary motor areas on the medial surface, not visible in Figure 4) and bilateral posterior inferior temporal cortex, substantial numbers of PtM-s and PtM-r electrodes were located around the posterior portion of the left Sylvian fissure in inferior parietal cortex and superior temporal cortex and near the anterior junction of the frontal and temporal lobes.

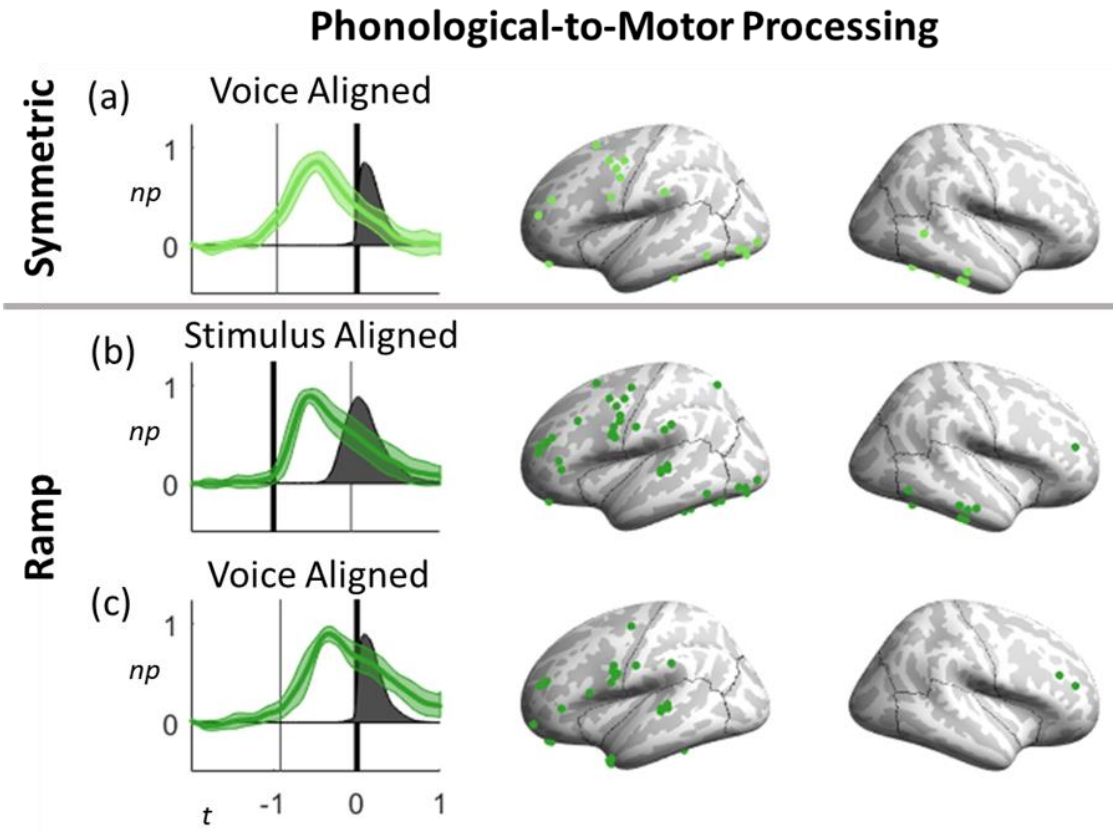


**Figure 4: Phonological-to-Motor Processing Clusters**

Two characteristic time courses included: (a) symmetric and (b, c) ramp. High gamma cluster activity patterns are shown on the left (normalized power (np) over time (t), in seconds), electrode locations for clusters on the right. Solid vertical lines show trial alignment condition and fainter vertical lines show average location of other alignment condition. Average audio signal amplitude indicated by gray shaded region. The symmetric pattern is only seen in the voicing alignment case (a), while the ramp activity is seen in the stimulus (b) and voicing (c) alignment cases.

### *III.4.iii Motor Execution*

Four clusters, falling into two characteristic time courses, had activity peaks shortly after voice onset. All four of these clusters exhibited symmetric time courses, with two showing a broader symmetric pattern and two a narrower pattern. Based on this timing and the observation that most of the electrodes exhibiting these time courses lie in primary sensorimotor and auditory cortical areas involved in speech articulation

(Guenther, 2016), these canonical time courses are termed *Motor Execution – Symmetric Broad (ME-sb)* and *Motor Execution – Symmetric Narrow (ME-sn).* The ME-sb canonical time course is shown in Figure 5a (stimulus aligned) and Figure 5b (voice aligned). The symmetric activity pattern shows a gradual ramp-up that initiates 80 ms after stimulus presentation and steadily increases until peaking 200 ms after voice onset, then gradually returns to baseline after production is complete. The two clusters exhibiting the ME-sn time course are shown in Figure 5c (stimulus alignment) and Figure 5d (voice alignment). Activity initializes 640 ms after stimulus presentation and peaks 110 ms after voice onset, with a narrower plateau than the ME-b time course. Brain areas with substantial numbers of electrodes exhibiting the ME-sb and ME-sn time courses included bilateral frontal and parietal cortex surrounding the ventral central sulcus, left inferior frontal cortex, right anterior frontal cortex, insula, superior and middle temporal cortex, and, to a lesser degree, inferior temporal cortex (see Figure 5).

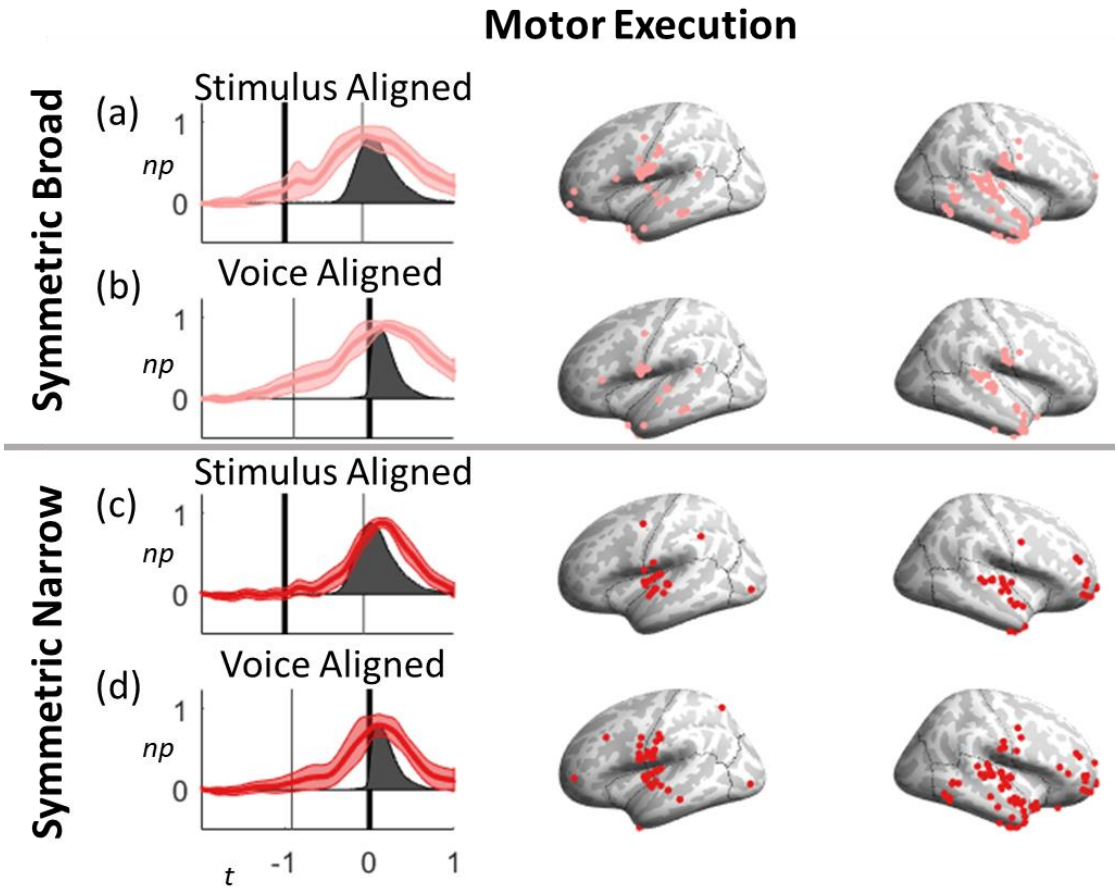


**Figure 5: Motor Execution Clusters**

Two characteristic time courses included: (a,b) symmetric broad and (c,d) symmetric narrow. High gamma cluster activity patterns are shown on the left (normalized power (np) over time (t), in seconds), electrode locations for clusters on the right. Solid vertical line show trial alignment condition and fainter vertical line show average location of other alignment condition. Average audio signal amplitude indicated by gray shaded region. Both broad and narrow clusters are seen for both stimulus alignment (a, c) and voicing alignment (b, d) cases.

### *III.4.iv Auditory Processing*

The final set of clusters exhibited later initiation of activity and fall primarily in the superior temporal gyrus and neighboring cortical regions (Figure 6). A symmetric activity pattern, termed the *Auditory Processing – Symmetric (AP-s)* time course, was exhibited by one cluster in each alignment case, as shown in Figure 6a (stimulus aligned)

and Figure 6b (voice aligned). In these clusters, activity begins to ramp up 10 ms after voice onset and is approximately centered around the self-produced acoustic signal, with a peak 570 ms after voice onset. A second time course, termed *Auditory Processing – Ramp (AP-r)*, was exhibited by one cluster in the voice aligned analysis (Figure 6c). In this cluster, activity quickly ramps up just prior to voice onset (initializing 50 ms before voicing) and peaks 200 ms after voice onset. This is followed by a gradual decay to baseline after production is complete.

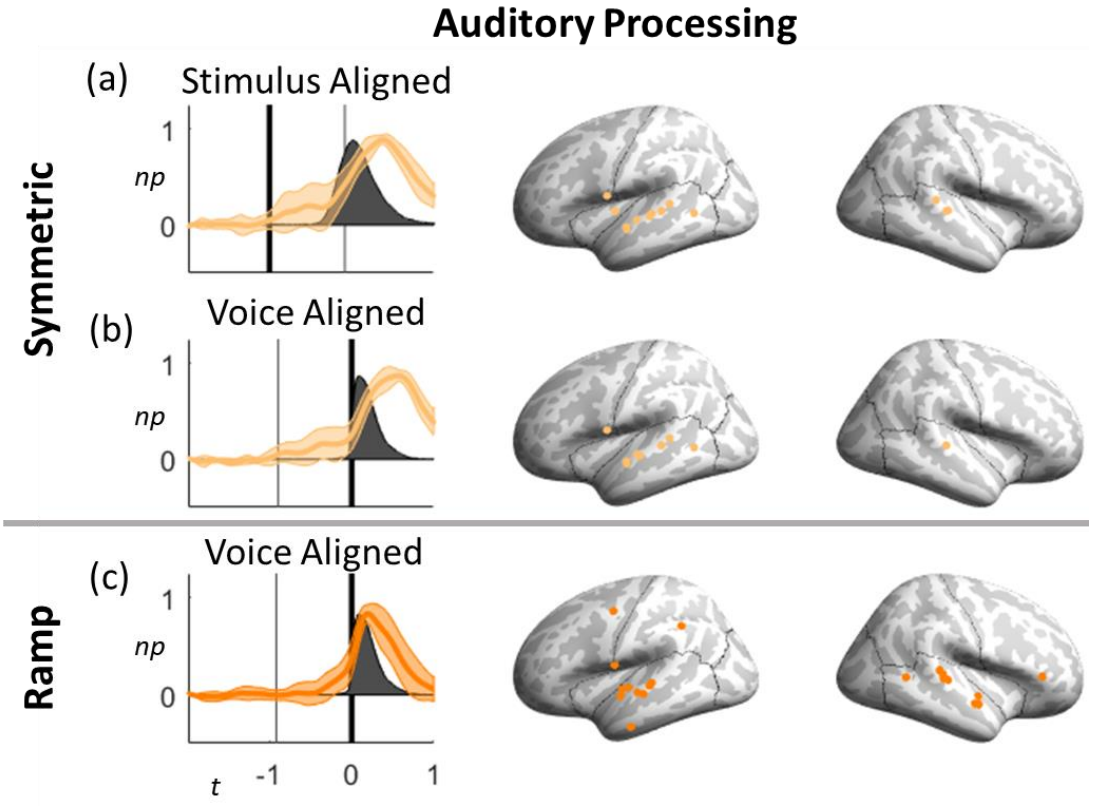


**Figure 6: Auditory Processing Clusters**

Two characteristic time courses included: (a,b) symmetric and (c) ramp. High gamma cluster activity patterns are shown on the left (normalized power (np) over time (t), in seconds), electrode locations for clusters on the right. Solid vertical line show trial alignment condition and fainter vertical line show average location of other alignment condition. Average audio signal amplitude indicated by gray shaded region. The symmetric pattern is seen in the stimulus presentation (a) and voicing onset (b) alignment cases, while the ramp activity is only seen in the voice aligned case (c).

## III.5 Discussion

In this study, a Kalman filter-based trend analysis and an unsupervised hybrid clustering procedure that combines partitioning and hierarchical clustering were used to identify eight canonical time courses of neural activity (as measured by signal power in the 70-150 Hz range) during production of simple speech utterances. Unsupervised clustering analysis takes a data-driven approach to segment responses into groupings with common representations while relaxing reliance on prior knowledge. The resulting clusters fell into four pairs, with each pair showing peak activity during one of four major stages of overt word reading: early stimulus processing, phonological-to-motor processing, motor execution, and auditory processing of the self-produced utterance. Furthermore, the large majority of electrodes displaying these canonical activity patterns were located in brain regions that have been associated, in prior studies, with the processing stages during which peak activity occurred. This pattern occurred in the absence of any prior knowledge of processing in the speech network, indicating that the analyses successfully extracted canonical time courses that reflect different functional components of the speech production process.

The identified clusters fell into two main shape classes: symmetric, in which activity increases from baseline at approximately the same rate it decreases back down to baseline, and ramp[5], in which activity increases quickly but ramps down gradually over

[5] Notably, the ramp pattern identified here is in a downward direction starting after the activity peak, whereas the ramp activity pattern identified by (Cheney and Fetz, 1980) from single unit motor cortex recordings during a manual isometric task was in an upward direction starting from baseline. This difference indicates that the ramp profile identified herein is not a human homolog of the ramp activity profile identified by (Cheney and Fetz, 1980) during primate movements.

the course of word production. The symmetric pattern occurred in all four processing stages, whereas the ramp pattern occurred in all but the motor execution stage. In terms of neural processing, the ramp profile indicates that the processing load for neurons in the region of the electrode rises quickly but only gradually decays until processing in that area is complete, and a symmetric activity pattern is indicative of a brain region where processing returns to baseline at the same rate that it arises.

Two of the eight canonical clusters, ESP-s and ESP-r, showed activity onsets within 40 ms of visual stimulus presentation and peaked within 210 ms. The ESP-s cluster then quickly returned to a near-baseline value within 310 ms (Figure 3a). Electrodes with this activity pattern were found in the left hemisphere near the junction of the temporal pole and lateral frontal orbital cortex. Lateral frontal orbital cortex has been associated with visual object identification, with higher activity for more confidently identified objects (Chaumon et al., 2014). In the right hemisphere, the ESP-s time course was found in middle and inferior posterior temporal cortex (PTC), regions which have been associated with saccadic eye movements (Zhou and Shu, 2017) and reading of non-word orthographic stimuli (Vigneau et al., 2005) such as those in our protocol. Although ESP-s activity returned nearly to baseline well before onset of voicing, it maintained a small amount of activity until the end of the trial (after vocalization was complete). In the cued auditory word repetition study, (Leonard et al., 2019) identified an activation time course exhibiting a similar pattern (cluster 3 in Figure 4 of (Leonard et al., 2019)); the authors suggest that this cluster may be involved in holding the word in working memory

during the delay period. Although our protocol does not involve a delay period, it may still require retention of phonological information until the end of the production process.

In contrast to the ESP-s cluster, the ESP-r cluster shows a more gradual activity decrease toward baseline, reaching its baseline value at approximately the same time that vocal output starts (Figure 3b). Like the ESP-s cluster, the ESP-r cluster was found in middle/inferior PTC regions associated with reading (bilaterally in this case). Additionally, an electrode with the ESP-r time course was found in left inferior frontal cortex near the junction of the posterior inferior frontal sulcus and the precentral sulcus, a region associated with speech motor programming (Guenther, 2016). The activity offset ramp of the ESP-r cluster overlaps substantially with activity in the clusters from the next processing stage, phonological-to-motor processing, with the PtM-s and PtM-r clusters reaching peak activity during the ramp-down of activity in the ESP-r cluster. The activity offset rate of the ESP-r cluster is very similar to the onset rates for the PtM-r and PtM-s clusters (see Table 2), suggestive of a gradual transfer of processing from the ESP-r electrode locations to the PtM electrode locations. This interpretation receives support from the ECoG study of (Collard et al., 2016), who identified significant high gamma interactions between inferior temporal electrodes (the general location of our ESP-r electrodes) and left inferior frontal electrodes (where a large proportion of our PtM-r and PtM-s electrodes are located) during the period between stimulus onset and vocal production in a picture naming task (see also (Korzeniewska et al., 2011)).

The next group of canonical time courses, the phonological-to-motor processing group, show peak activity approximately 250-300 ms after the early stimulus processing

group and 340-480 ms before voicing onset (Figure 4). In addition to the anatomical locations exhibiting the ESP-r and ESP-s time courses, the PtM-s and PtM-r time courses are particularly prevalent in left hemisphere frontal cortex, with only a handful of electrodes in right frontal cortex. Particularly dense concentrations are found around left inferior frontal sulcus, an area associated with phonological working memory (Buchsbaum et al., 2005; Rottschy et al., 2012), and left ventral premotor cortex, an area associated with speech motor planning (Guenther et al., 2006; Guenther, 2016). Left lateralization of the PtM clusters is in keeping with the well-known association between left inferior frontal cortex and language output processing (e.g., (Broca, 1861, 1865; Friederici, 2017)) as well as speech motor planning (Flinker et al., 2015; Guenther, 2016), supporting the interpretation that these clusters are involved in phonological processing, including translation of phonological information into speech motor programs. In this regard, the gradual offset of the PtM-r traces, which continues through the production period, may be indicative of a gradual decrease in working memory and speech motor programming load as each phoneme in the utterance is generated. This view is supported by the fact that the offset rate of the PtM-r cluster is approximately the same as the onset rate of the motor execution cluster ME-sb, which is suggestive of a gradual hand-off of processing from planning to execution mechanisms starting roughly 350 ms before voice onset and continuing until articulation is near completion. Support for this view comes from (Korzeniewska et al., 2011), who found directed causal interactions between a site in left inferior frontal gyrus (where many PtM-r electrodes are found) to a number of sites in perisylvian areas and ventral sensorimotor cortex (where

many electrodes exhibiting canonical motor time courses are found in the current study) prior to spoken responses in a picture naming task and an auditory word repetition task.

In addition to the locations described above, PtM-s and PtM-r activity is also found in left hemisphere auditory and somatosensory cortical regions near the posterior portion of the Sylvian fissure. This activity may be related to the retrieval of phonological codes (Buchsbaum et al., 2001) and/or the generation of auditory and somatosensory expectations for upcoming speech sounds (Guenther, 2016).

The motor execution group consisted of broad and narrow canonical time courses that peaked shortly after voice onset, both with symmetrical shapes (Figure 5). A substantial percentage of electrodes from this group surrounded the central sulcus in the ventral sensorimotor cortex, where motor and somatosensory representations of the speech articulators are located (Penfield and Roberts, 1959; Takai et al., 2010; Bouchard et al., 2013; Guenther, 2016). Activity onset in the ME-sn time course begins 380 ms prior to voicing onset, with activity returning to baseline at approximately the same time or shortly after the acoustic signal ends. Given that the delay between motor cortical activity and movement onset is approximately 40 ms (Guenther et al., 2006) and articulatory movements for an isolated utterance can begin hundreds of ms prior to vocal onset (Conant et al., 2018), it is likely that the ME-sn time course is driven at least in part by motor cortical neurons initiating movements of the vocal tract musculature, along with corresponding somatosensory cortical responses. Notably, however, electrodes displaying the ME-sn activity pattern are not limited to sensorimotor cortex; a large number of ME-sn responses were also found in superior temporal cortex and inferior frontal cortex, both

with a right hemisphere bias. These regions have been associated with auditory feedback control mechanisms in speech production that monitor auditory feedback of one's own speech and generate motor commands to correct perceived deviations from the desired acoustic signal (Guenther et al., 2006; Guenther, 2016).

The ME-sb time course was found in most of the same cortical regions as the ME-sn time course with the exception of right inferior frontal regions; instead some electrodes exhibiting the ME-sb pattern were found in left inferior frontal cortex. Activity in the ME-sb cluster begins much earlier than in the ME-sn cluster, with an onset time more than 700 ms prior to the onset of vocal output. This is suggestive of motor planning processes, consistent with the view that left inferior frontal cortex is more involved in the generation of speech motor programs in a feedforward fashion, whereas right inferior frontal cortex is more involved in sensory feedback-based control mechanisms (Guenther et al., 2006). Activity in the ME-sb cluster does not return to baseline until well after vocalization is complete. Similarly, (Leonard et al., 2019) similarly found a cluster throughout much of sensorimotor and auditory cortex whose activity started more than 1 second prior to vocal output, peaked approximately 200 ms after vocal onset, and extended beyond the acoustic output. The late return to baseline, well after the acoustic signal has ended, seems, on the surface, to be inconsistent with a motor planning interpretation. Notably, however, articulator movements in isolated utterances can continue hundreds of ms beyond vocal offset (Conant et al., 2018), suggesting that movement may still be ongoing when the ME-sb cluster returns to baseline.

The final two canonical time courses, constituting the auditory processing group (Figure 6), had activity onsets that occurred near voice onset and activity peaks that occurred after the peak of the sound envelope of the vocalized utterance. Activity in the AP-r cluster increased rapidly with a more gradual offset of activity. The AP-s cluster showed a similar activity offset rate but a more gradual onset, resulting in its symmetric shape. The offset rate of both AP clusters approximately follows the offset rate of the acoustic envelope (see Figure 6). Given that activity in these clusters lags the vocal output of the subject and that the corresponding electrodes are predominantly located in superior temporal cortex, the AP-r and AP-s clusters very likely represent auditory cortical responses to one's own voice during speech.

In summary, our results indicate that neural activity underlying the production of orthographically presented syllables falls broadly into two temporal profile shape categories, ramped and symmetric. Furthermore, distinct characteristic time courses are found during four different task stages: early stimulus processing, phonological-to-motor processing, motor execution, and auditory processing of self-produced speech, with activity offset ramps in earlier stages approximately matching activity onset rates in later stages. Finally, the analysis tools developed in the current study, most notably the Kalman filter-based trend analysis, provide a powerful means for identifying and quantitatively characterizing the neural computations underlying human cognition and behavior beyond the domain of speech production.

## III.6 Supplemental

### *III.6.i Comparison Between Alignment Conditions*

Two alignment conditions were considered in the analysis of this chapter and presented in the results in §III.4, alignment by the presentation of the orthographic stimulus and alignment by the onset of voicing. The results and discussion, §III.4 and §III.5, discussed a set of canonical time courses across both alignment conditions, with a total of eight characteristic temporal profiles. It was noted that four of the eight were present in both alignment conditions. Here we justify using the results of the two alignment conditions as a combined set of characteristic time courses instead of breaking them up into two separate and distinct sets of results.

The four temporal profiles that were present in both alignment cases were Phonological-to-Motor ramp (PtM-r), Motor Execution symmetric broad (ME-sb), Motor Execution symmetric narrow (ME-sn), and Auditory Processing symmetric (AP-s). The average voicing onset occurred 916 ms after stimulus presentation, as noted in §III.3.vii. Using this, we plot the time course of two alignment cases for each of the four cluster on top of each other in Figure 7, with the stimulus alignment case plotted with solid lines and the voicing onset plotted with dashed lines. The figure is broken up by cluster: (a) PtM-r, (b) ME-sb, (c) ME-sn, and (d) AP-s. Further, each cluster has two plots: (-i) aligning the time course for the two alignment conditions by the average delay in voicing onset, e.g., shifting the voicing onset time course by 916 ms, and (-ii) aligning the time course for the two conditions by the location of their peak values.

It is seen that there is a high degree of similarity between the activity patterns collected from the two alignment conditions, giving strong evidence that the analysis was correct in combining results across to the conditions and further that these are indeed characteristic time courses. The similarity is strong when using the average latency in voicing onset (Figure 7 (-i)) and is seen to only have minor fluctuations when aligned by the peak values (Figure 7 (-ii)).

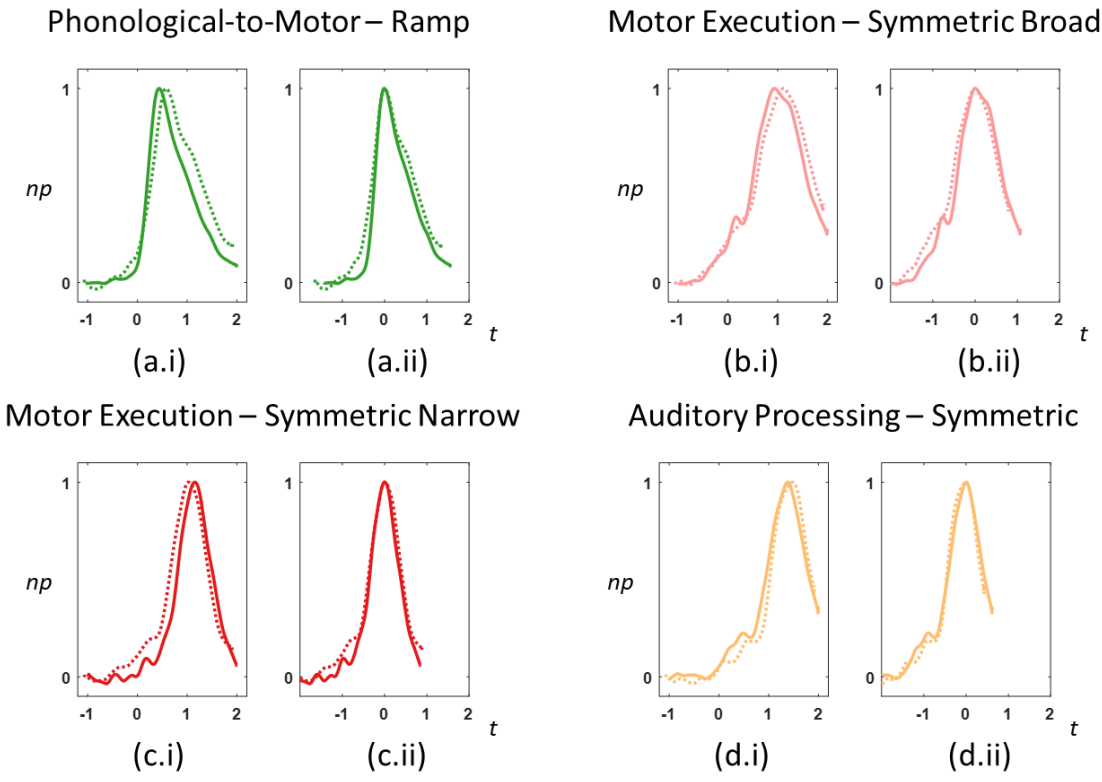


**Figure 7: Comparison between Clusters Seen in Both Alignment Conditions**

The four clusters seen in both the stimulus presentation (solid lines) and voicing onset (dashed lines) alignment conditions are plotted together for an illustration of the visual similarity: (a) PtM-r, (b) ME-sb, (c) ME-sn, and (d) AP-s. Normalized power (np) is plotted over time (t). In (-i) t=0 represents the time of the stimulus presentation. Voicing onset time courses have been shifted by 916 ms (average voicing onset latency) to align the responses from the two cases. In (-ii), both time courses have been shifted to have their peaks located at t=0. Cluster colors have been maintained from previous figures for completeness.

### *III.6.ii Gradual Transfers of Processing*

In §III.5 it was commented that the offset of earlier processing stages overlaps with the activity onset in the next processing stage, with their activity rates, onset and offset, having a high degree of similarity. This was attributed to a gradual transfer of processing from the earlier processing stage to the next stage. Here we explore this observation more, highlighting the power of our methodology which allows for quantitative assessment of trends.

Clusters in all four processing stages were present in the stimulus alignment presentation case. Inspecting Table 2, it can be clearly seen that there are cases where an earlier processing stage offset rate matches closely with the onset rate of the next processing stage. To quantitatively measure rates that match closely, the same threshold is used as was used for defining a symmetrical shape, i.e., the absolute value of the offset rate needs to be within 10% of the value of the onset rate (refer to §III.3.ix). This is found in three clear cases: 1) from Early Stimulus Processing ramp (ESP-r) to Phonological-to-Motor ramp (PtM-r), 2) PtM-r to Motor Execution symmetric broad (ME-sb), and 3) Motor Execution symmetric narrow (ME-sn) to Auditory Processing symmetric (AP-s).

These results are further illustrated in Figure 8, with (a) for ESP-r to PtM-r, (b) for PtM-r to ME-sb, and (c) for ME-sn to AP-s. In Figure 8 each pair of processing stages is shown in subplot (-.i), i.e., (a.i) for ESP-r to PtM-r. Subplots (-.ii) zoom in on the part of the trial period between the peak of the activity in the earlier and later processing stage. Visual inspection of this figure shows that the activity of the earlier

processing stage is decreasing at the time that activity in the next stage is increasing, and at about the same absolute rate. To make this more explicitly striking, subplot (-.iii) provides a view where the offset rate of the earlier processing stage is flipped to show a positive trend and aligned with the peak of the next processing stage. For example, (a.iii) shows the onset rate for PtM-r in green, displaying the same result as in (a.ii), but ESP-r, the earlier processing step in the pair, has been flipped so that the negative offset rate in (a.ii) appears as a positive rate in (a.iii). It is then shifted so that the peaks align. This is done so that the offset and onset rates between the two stages can viewed in direct comparison. Visually inspecting the results across the three pairs of results show a very high visual alignment between earlier stage offset rates with the next stage onset rates, providing further evidence of the potential for a gradual transfer of processing from one stage to the next, and from one cluster to another.

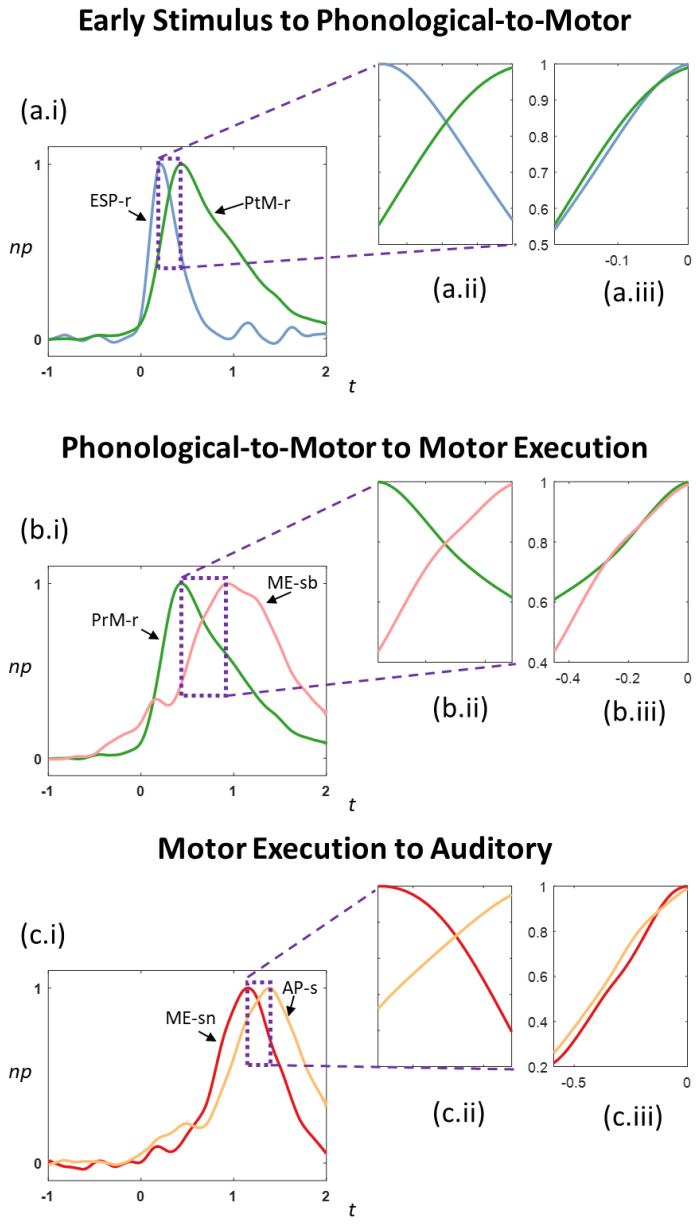


**Figure 8: Gradual Transfers of Processing in Stimulus Alignment Case**

Normalized power (np) over trial time (t), in seconds, for paired processing stages: (a) Early Stimulus Processing ramp (ESP-r) to Phonological-to-Motor ramp (PtM-r), (b) PtM-r to Motor Execution symmetric broad (ME-sb), and (c) Motor Execution symmetric narrow (ME-sn) to Auditory Processing symmetric (AP-s). Each processing pair contains three plots: (-.i) entire trial period, (-.ii) zoomed in to period between activity peaks where earlier processing stage offset rate and next processing stage onset rate can be seen, (-.iii) similar view to (-.ii) but flipped (over y-axis) version of earlier processing stage with peaks aligned. All cluster colors maintained to match other figures in this chapter.

Figure 9 shows the same layout, but for the onset alignment case. In this alignment case the Early Stimulus Processing stage was not present. Three pairs of processing stages exhibited this gradual transfer of processing: 1) Phonological-to-Motor symmetric (PtM-s) to ME-sn, 2) PtM-r to ME-sb, and 3) ME-sn to AP-s, as shown in Figure 9 (a), (b), and (c), respectively. The plots are laid out similar to Figure 8, with numerical results presented in Table 2.

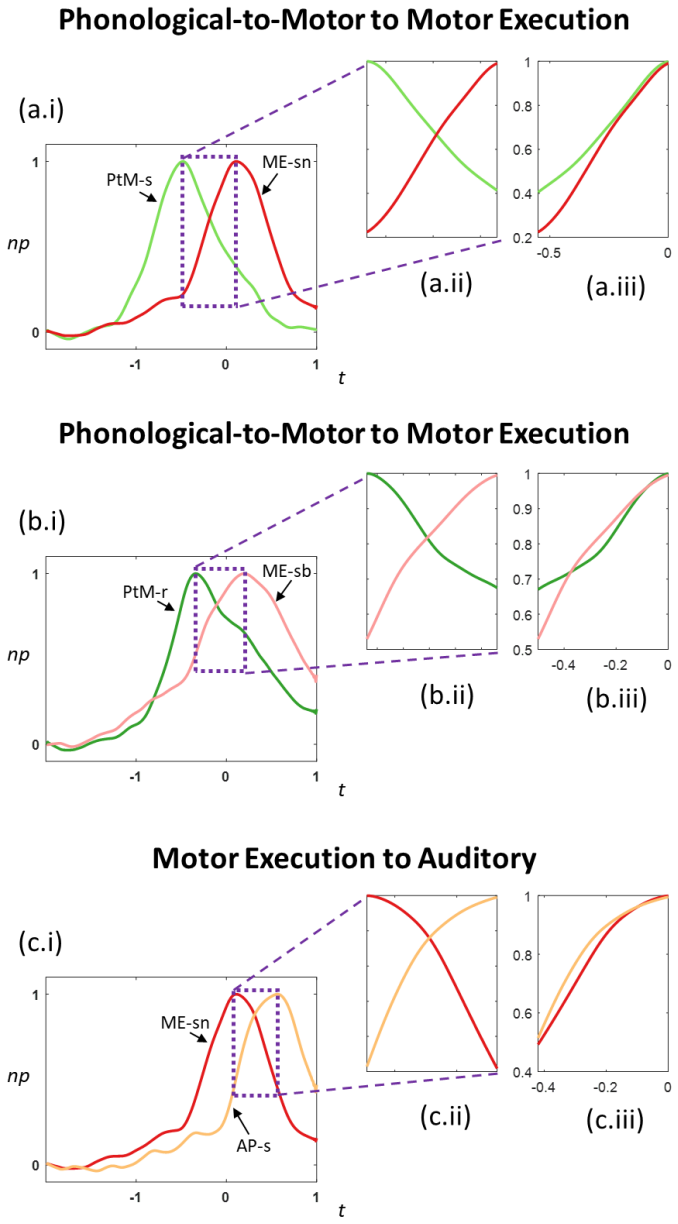


**Figure 9: Gradual Transfers of Processing in Voicing Alignment Case**

Normalized power (np) over trial time (t), in seconds, for paired processing stages: (a) Phonological-to-Motor symmetric (PtM-s) to Motor Execution symmetric narrow (ME-sn), (b) Phonological-to-Motor ramp (PtM-r) to Motor Execution symmetric broad (ME-sb), and (c) ME-sn to Auditory Processing symmetric (AP-s). Each processing pair contains three plots: (-.i) entire trial period, (-.ii) zoomed in to period between activity peaks where earlier processing stage offset rate and next processing stage onset rate can be see, (-.iii) similar view to (-.ii) but flipped (over y-axis) version of earlier processing stage with peaks aligned. All cluster colors maintained to match other figures in this chapter.

Gradual transfers from PtM-r to ME-sb and ME-sn to AP-s were seen in both alignment cases, while ESP-r to PtM-r and PtM-s to ME-sb were only seen in the stimulus presentation and voicing onset alignment cases, respectively. The ESP-r cluster was only present in the stimulus presentation alignment case, while the PtM-s cluster was only present in the voicing onset cluster.

### *III.6.iii Shape Similarity Across Processing Stages*

Another type of analysis that falls out of being able to quantitatively characterize trends is the ability to see if the same activity pattern is present in different processing stages. Assessing similarity in the trends between change points allows quantitative measurement of activity shape similarity. This is conducted by measuring the similarity in the onset rate, offset rate, and plateau duration (if it exists). Again, a 10% threshold is used to define similarity, as was discussed in §III.3.ix.

Phonological-to-Motor symmetric (PtM-s) and Motor Execution symmetric narrow (ME-sn) are found to have the same activity pattern in the voicing onset alignment case, which was the only case that has PtM-s, as can be quantitatively seen in Table 2. If the similarity threshold is relaxed from 10% to 15%, Auditory Processing symmetric (AP-s) is seen to have the same activity as PtM-s in the voicing onset case and with ME-sn in both alignment cases (stimulus presentation and voicing onset).

Figure 10 illustrates this finding. Figure 10 (a) displays PtM-s, ME-sn, and AP-s for the voicing onset alignment condition. Figure 10 (b) shows the three clusters when aligned by their peaks, illustrating the strong similarity in their processing shapes. The peak of PtM-s precedes ME-n by 590 ms and AP-s by 1050 ms. ME-n precedes AP-s by

460 ms. Close inspection of Figure 10 (b) shows AP-s slightly deviating from the other two clusters during the onset rate, which is why it does not fit a similarity score of 10%, but does at 15%.

Additional inspection of Figure 10 (b) presents another interesting result. The average audio signal amplitude, indicated by the gray shaded region, appears to show a relatively similar offset duration as the offset durations of the three clusters. We did not dig into this further, but make this observation in the hope that others may explore further.

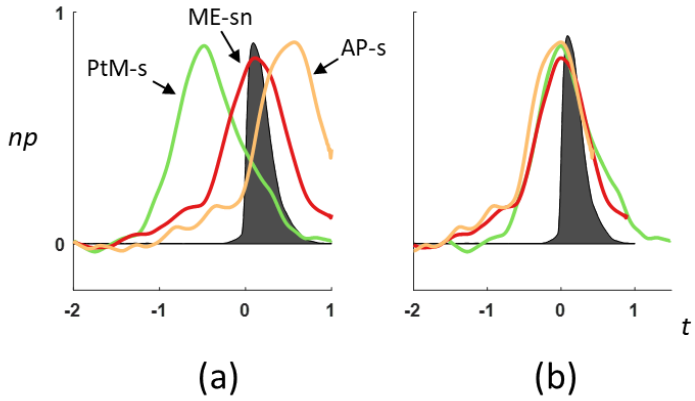


**Figure 10: Same Activity Across Processing Stages**

Normalized power (np) over time (t), in seconds, for clusters from voicing onset alignment case with same activity pattern. (a) Phonological-to-Motor symmetric (PtM-s), Motor Execution symmetric narrow (ME-sn), and Auditory Processing symmetric (AP-s) clusters plotted relative to trial time. (b) PtM-s, ME-sn, and AP-s plotted when aligned by their peak value. Average audio signal amplitude indicated by gray shaded region. All cluster colors are maintained throughout the chapter.

### *III.6.iv Same Offset, Different Processing Stage*

Quantitative assessment of trend comparisons can also be done in absolute terms. In §III.6.ii and §III.6.iii trends were compared in relative terms, i.e., by shifting the signals (and in §III.6.ii flipping them) to have them overlap in time. In contrast, comparisons in absolute terms look for not only for the trends to be the same, but also for

the time they occur in the trial to be the same. The same threshold of 10%, §III.3.ix, is used for rate similarity, but with the addition of needing to adhere to absolute temporal location of where the rate occurs.

The criteria for this additional analysis was met between a pair of clusters in the voicing onset alignment condition – Motor Execution symmetric narrow (ME-sn) and Auditory Processing ramp (AP-r). ME-sn and AP-r were found to have the same offset rate and absolute timing within the trial for the entirety of their offsets, as shown in Figure 11 (c). This opens up potential future areas of research to explore why activity from separate processing steps are deactivating at the same time and at the same rate.

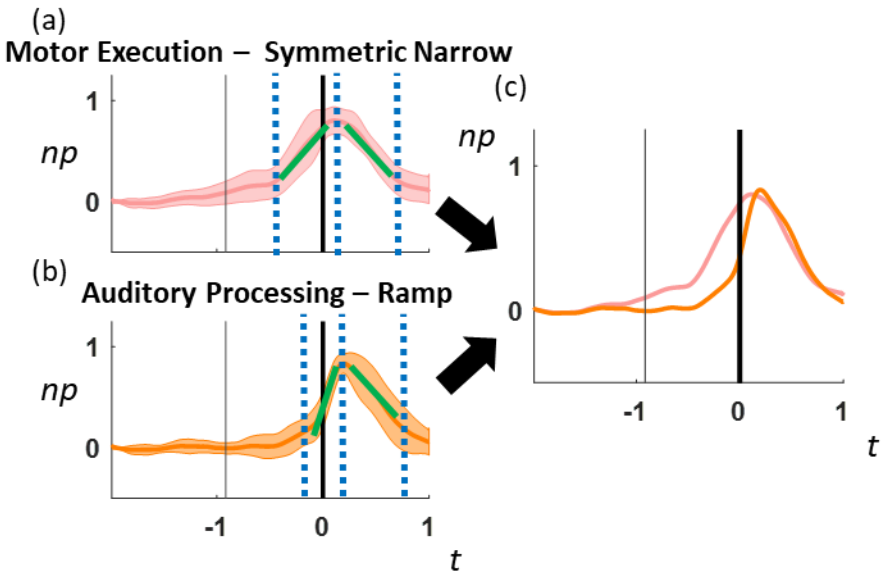


**Figure 11: Identical Offset Activity from Different Processing Stages**

Normalized power (np) over time (t), in seconds, for (a) Motor Execution symmetric narrow (ME-sn) and (b) Auditory Processing ramp (AP-r). (c) Clusters plotted together to illustrate same offset activity.

# CHAPTER IV: Anatomically Constrained Clusters

## IV.1 Introduction

Functional mappings of the brain have resulted in distinct regions for various functionalities, with neighboring regions for processing similar functions, i.e., motor processing of the index finger is anatomically co-located and next to processing for the middle finger (Penfield and Boldrey, 1937; Penfield and Roberts, 1959). This has created anatomical parcellations that are also functional parcellations. The speech network has been previously shown to have a functional breakdown across anatomical regions of the brain, starting with lesion studies that identified Broca's and Wernicke's areas (Broca, 1861, 1865; Wernicke, 1874) and continuing with numerous studies since, see (Geschwind, 1979; Guenther, 2016) for a review.

In CHAPTER III, characteristic time courses during speech were found in the absence of any assumptions or restrictions on the anatomical locations of the recording electrodes. This approach let the underlying data empirically drive the findings without prior knowledge of the anatomical and functional brain network. Results showed diffuse activity across many brain regions, with a cluster functional breakdown that matches speech processing flow (Guenther, 2016), namely Early Stimulus Processing, Phonological-to-Motor Processing, Motor Execution, and Auditory Processing. It was seen, and discussed, that activity within these functionally distinct processing clusters also had anatomical preferences. Several clusters had diffuse activity across brain regions, but were seen to have anatomical focal areas.

In this chapter we explore how results change when anatomical constraints are included to enforce locality within brain regions. Our goal is to tease out and separate the observed focal areas found in CHAPTER III and see if any different results emerge. Anatomical locality is enforced in the creation of clusters, biasing results towards having both local anatomical and functional (temporal profile) similarity. We find that the results generally still hold, with sub-clusters forming in each of the clusters previously found that are more anatomically constrained to the observed focal areas. We first discuss how the approach is updated to enforce an anatomical bias in the cluster creation. We then present the results and a discussion before concluding.

## IV.2 Methods

The exponentially weighted distance measure from CHAPTER III, Equation 1, is modified to encourage anatomical locality. A spatial kernel (Kubanek and Schalk, 2015) is used to provide the constraint by setting an anatomical probabilistic distribution over which local anatomical regions could have similar functionality. Spatial kernels are designed to give differential weight in a spatial pattern so some regions contribute more to the result while others contribute less. In this application we want regions close to an electrode to get highly weighted as being anatomically similar, while those further away get less weight. Thus, the kernel should produce a high value for two electrodes close together and a low value to two electrodes that are far away, such as the frontal lobe and occipital lobe.

Unless otherwise specified, all the methods from CHAPTER III are applied to this analysis as well. Additional details on the modifications to the methods beyond what is discussed in the following subsections can be found in APPENDIX B.

### *IV.2.i Anatomical Representation*

Anatomical space is measured using the Montreal neurological institute (MNI) coordinates. Electrode locations are provided in MNI coordinates as covered in CHAPTER III. Therefore, anatomical distances will be measured as differences in the MNI coordinates between electrodes.

It is desired to allow bilateral clusters to form and not unnecessarily limit resulting clusters to be unilateral. Bilateral activations in similar anatomical regions have been found during speech processing stages, i.e., auditory reception in superior temporal gyrus (STG) (Guenther, 2016). The absolute value of the MNI X-coordinate, which captures the lateral displacement, is taken to allow for the creation of both unilateral and bilateral clusters.

Analysis was also conducted with the raw value of the X-coordinate. Results from this additional analysis were similar to those discussed later in this chapter, but the bilateral clusters were broken up into two clusters, one for each hemisphere. These results are not shown or discussed further.

### *IV.2.ii Spatial Kernel*

The spatial kernel used in this analysis takes the form of a Gaussian radial basis function (RBF) with a plateau of similarity near the center and then a Gaussian roll-off.

This provides a high measure of similarity for electrodes that are close to one another, a region of roll-off for decreasing similarity as distance is increased, and near-zero values for electrodes are very far away. These are desired properties, yet are manifested in a simple kernel allowing for minimal parameterization.

A radial spatial relationship, however, does not accurately capture the anatomical construct of the brain, as such dynamics as cortex folding, i.e., sulci and gyri, create spatial differences that the RBF does not take into account. Other kernels could be used here (Potworowski et al., 2012; Chintaluri and Wójcik, 2015), but applying a RBF kernel was found to provide enough insight into the results of this analysis to allow conclusions to be drawn without greatly increasing complexity. A different kernel could have been used if a larger population of subjects was available and tighter anatomical clusters was desired. For example, a different type of kernel may be required if the desire is to anatomically break up different articulators for speech production. We leave this type of extension to others.

The RBF kernel is parameterized as follows. A locality factor, $\theta$, controls the degree of spatial locality that the kernel enforces. This locality factor is used to set the radius of plateau for perfect spatial similarity and also for the shape and roll-off rate for the kernel. In the final implementation, $\theta$ was used for the plateau radius and both the mean and standard deviation of the Gaussian roll-off, as shown in Equation 7. After analysis with empirical data, the locality factor was set to *20* based the anatomical coverage it provided. This resulting coverage had a general spatial extent on the order of coarse speech regions, i.e., Wernicke's area. This was the goal of the spatial kernel, to

cover coarse anatomical regions as opposed to fine regions. Early analysis explored different values for $\theta$, providing coarser or finer spatial extent. It was found that many regions lacked enough electrodes to support finer regions, while coarser regions got too large and resulted in similar findings to CHAPTER III. This value was not optimized, and similar to the kernel used, more time could have been devoted to optimizing this value if this was the main focus of the research. Future research could explore optimizing this value more if it is required for other purposes.

Equation 7 shows the RBF kernel used for enforcing spatial locality between electrodes, $kernel(\Delta)$. The Euclidean distance between the MNI coordinates of a pair of electrodes, *A* and *B*, is taken as the distance $\Delta$. The vectors *A* and *B* are the MNI coordinates of the two electrodes that are being compared, with the absolute value of the X-coordinate being used to enable bilateral clusters, as described in §IV.2.i. The kernel then operates on $\Delta$ to compute spatially similarity. The kernel is scaled by the locality factor just described, $\theta$. The scale factor, $\beta$, is grouped with the mixing factor for convenience, as described in §IV.2.iii. The maximum in the exponent serves to create a plateau with radius $\theta$ where there is no difference in spatial distance.

**Equation 7: Spatial Kernel**

$$kernel(\Delta) = \beta\, e^{-\frac{(\max(\Delta,\ \theta)-\theta)^2}{2\theta^2}}, \qquad where\ \Delta = \sqrt{\sum (A-B)^2}$$

*IV.2.iii Distance Measure*

The weighted spatial similarity of two electrodes from Equation 7 is then combined with the temporal difference between the electrodes' high gamma power to

form the distance measure for clustering. The temporal difference is the distance measure as described in Equation 1 of CHAPTER III (§III.3.viii), here referred to as *TempDiff*. To ensure the two distances (the spatial and temporal) are scaled similarly, the logarithm of the temporal distance is taken. Doing so requires adding one to *TempDiff* to enforce all values to be finite. Several different ways to combine the two distances were researched, but the best performance was found when using a mixing term[6]. Best performance was measured empirically through trial and error by comparing resulting temporal and spatial similarities for different mixing factors. The combined distance measure is provided in Equation 8. The mixing term, i.e., the product of the two distance measures, is scaled by a factor, $\alpha$, and added to the expression for the temporal difference term, *TempDiff*. The scaling factor $\alpha$ subsumes $\beta$ (Equation 7) for convenience and was set experimentally.

**Equation 8: Distance Measure with Anatomical Constraint**

$$distance(A,B) = \ln(TempDiff + 1) * (1 + \alpha\, kernel(\Delta))$$

All other methods are the same as in CHAPTER III. The distance measure of Equation 8 replaces Equation 1, with the rest of the steps following those described in CHAPTER III.

[6] Alternative ways to combine the temporal and spatial distance measures were explored and are discussed in APPENDIX B.

## IV.3 Results and Discussion

This analysis creates a degree of variability in the results and thus the discussion on the results will be limited. Results generally support the findings of CHAPTER III, primarily the breakdown by characteristic activity patterns, i.e., ramp and symmetric, and functional segmentation into four processing groupings: Early Stimulus Processing, Phonological-to-Motor processing, Motor Execution, and Auditory Processing.

The variability is primarily driven by the small number of electrodes that fall into several of the resulting clusters, stemming from the small subject sample size, and the mixing factor, $\alpha$, in Equation 8. This mixing factor tries to balance temporal similarity with spatial locality of the electrodes. If it weights too much in favor of the spatial location, resulting clusters become tight anatomical parcellations with time courses that have large standard deviations and do not accurately capture the temporal profiles of individual electrodes in the cluster. Going too far the other way results in the temporal similarity dominating the resulting clusters, with clusters very similar to CHAPTER III. Finding the right balance between these two is complicated by the limited number of electrodes. The spatial kernel cannot be reduced too small with the limited number of clusters, as the resulting clusters will each have very few electrodes in them. Thus clusters must represent more coarse spatial regions, i.e., on the order of Wernicke's area. This enabled the creation of clusters with enough supporting electrodes in them. Future research with more subjects should revisit the work of this section to explore the finer anatomical localization that can result if there are more supporting electrodes.

With these limitations, this section will be narrow in its scope. It will primarily serve to highlight and motivate the approach and potential benefit of this type of analysis, but to not put too much weight in the results. It shows promise, however, that results followed those of CHAPTER III, with the results of this chapter largely being an anatomical sub-division of CHAPTER III. This gives more support to the canonical temporal profiles found in the purely data-driven approach of CHAPTER III, and also helps support the characterizations found, the symmetric and ramp activity patterns.

For this discussion, we will only focus on the stimulus presentation alignment case. A total of 12 clusters were found using the change in distance measure to include spatial similarity, Equation 8. Results will be provided first, followed by a discussion connecting the results back to the findings generated without the additional anatomical constraint. Results naturally fell into the same four groupings as CHAPTER III: Early Stimulus Processing (ESP), Phonological-to-Motor (PtM) processing, Motor Execution (ME), and Auditory Processing (AP). Results will be presented in this order. Colors will be used to differentiate clusters within a group, e.g., within ESP, but are reused from group to group. We will refer to the results from this chapter as being anatomically constrained (AC) and will refer to the results from CHAPTER III as unconstrained or simply without the addition of AC in the prefix.

Anatomically constrained results from the voicing onset alignment case were the corollary of the onset alignment results of CHAPTER III and will not be discussed.

### *IV.3.i Early Stimulus Processing*

Two clusters were found in the early stimulus processing (ESP) grouping that showed activity that is consistent with processing the visual orthographic stimulus, as shown in Figure 12. Both clusters have a similar activity increase rate following stimulus presentation, followed by a decay back to baseline activity after a short duration of peak activity. The green cluster shows some prolonged activity just above baseline levels for the duration of the task. Both clusters show a slight increase in activity just after the peak of the acoustic response, which is only significant for the green cluster. Anatomically, both clusters are located in the posterior and ventral regions of the temporal lobe.

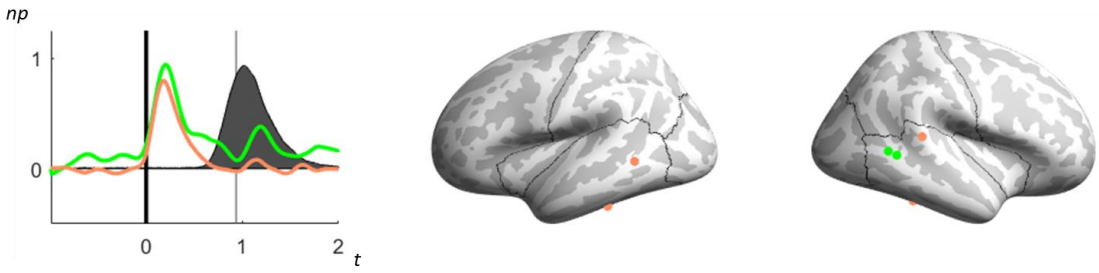


**Figure 12: Anatomically Constrained Early Stimulus Processing Clusters**

Left – Characteristic high gamma power (normalized power (np)) time courses (time (t) in seconds). Solid vertical line indicates stimulus presentation and fainter vertical line shows average location of voicing onset. Average audio signal amplitude indicated by gray shaded region. Right – electrode locations for clusters on the left.

Individual cluster results are broken out in Figure 13 (a) & (b). In Figure 13 (a), the green cluster has a symmetric shape until just prior to reaching baseline activity, where there is an activity tail as the power remains variable and just above baseline levels. We call this cluster Anatomically Constrained ESP symmetric (AC ESP-s). An additional significant, but small symmetrical peak in activity starts just after voicing

onset and ends as voicing concludes. Activity in this cluster is localized to middle posterior temporal cortex (PTC) in the right hemisphere. This region has been associated with saccadic eye movements (Zhou and Shu, 2017) and reading of non-word orthographic stimuli (Vigneau et al., 2005). The small elevated levels of activity throughout the duration of the task is similar to what (Leonard et al., 2019) found and attributed to working memory.

The orange cluster in Figure 13 (b) has a ramp activity pattern, with a quick ramp up to peak activity immediately following stimulus presentation and then a slower decay back to baseline activity. We call this cluster Anatomically Constrained ESP ramp (AC ESP-r). Larger variance is seen in the decay versus the ramp up. Bilateral activity was found in the inferior temporal cortex, with additional middle PTC activity in the left hemisphere and superior PTC activity in the right hemisphere, regions also found to be involved in saccadic eye movements (Zhou and Shu, 2017) and reading of non-word orthographic stimuli (Vigneau et al., 2005) and visual stimuli (Iacoboni and Dapretto, 2006; Buchweitz et al., 2009).

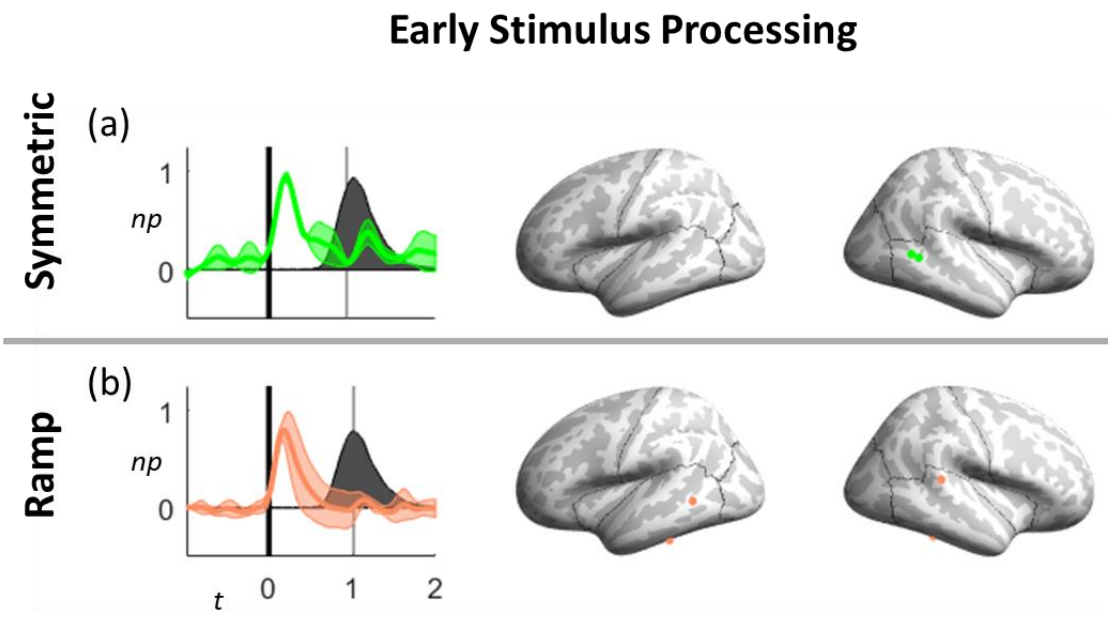


**Figure 13: Anatomically Constrained Early Stimulus Processing Individual Clusters**

Anatomically Constrained Early Stimulus Processing (AC ESP) results are broken into two characteristic time courses: (a) symmetric and (b) ramp. High gamma cluster activity patterns are shown on the left (normalized power (np) over time (t), in seconds), electrode locations for clusters on the right. Solid vertical lines indicate stimulus presentation and fainter vertical lines show average location of voicing onset. Average audio signal amplitude indicated by gray shaded region.

The anatomical constraint in this analysis resulted in two ESP clusters, each with limited number of electrodes present and constrained to the temporal lobe. Both symmetric and ramp activity patterns were seen, but higher variance in the activity patterns is also seen due to the limited number of electrodes.

### *IV.3.ii Phonological-to-Motor Processing*

Four clusters fit the functional profile of phonological-to-motor (PtM) processing, with activity largely in premotor and word form areas as shown in Figure 14. This grouping is primarily localized to the left hemisphere, with bilateral activity only present in one of the clusters and constrained to ventral regions, see orange electrodes in Figure

14. It is well known that the left hemisphere is involved with language processing (Broca, 1861, 1865; Friederici, 2017). In particular, it has been found to be specific for speech motor planning (Flinker et al., 2015; Guenther, 2016). This lateralization, along with the timing discussed next, help support this group being given the name phonological-to-motor processing.

The activity in this grouping was similar across the four clusters, showing a ramp activity pattern. The clusters were primarily broken up anatomically. In all clusters, activity starts just after stimulus presentation and peaks around half a second before voicing onset. Decay patterns showed large tails in most of the clusters, with activity not returning to pre-task baseline levels until after voicing is complete. Activity in the tails of the activity patterns is where the clusters differ the most, as seen in Figure 14. A large slowly decaying tail in the blue cluster separates that cluster from the rest, along with a slightly delayed peak. The red cluster is also distinct in its initial symmetric shape that turns into a slow decaying tail. The red, orange, and green clusters all have the same activity rise patterns, with the blue cluster lagging. Each cluster will now be discussed in a little more detail, with individual cluster break out results shown in Figure 15.

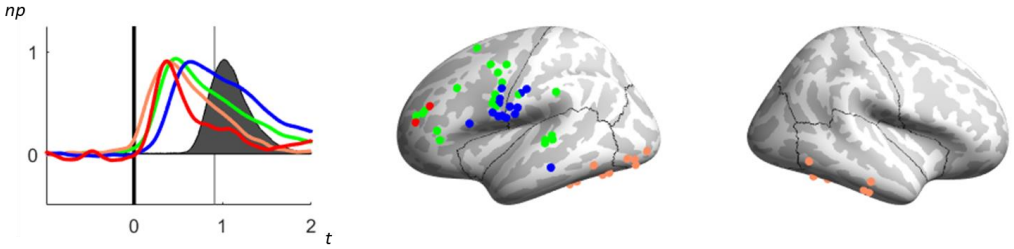


**Figure 14: Anatomically Constrained Phonological-to-Motor Processing Clusters**

Left – Characteristic high gamma power (normalized power (np)) time courses (time (t) in seconds). Solid vertical line indicates stimulus presentation and fainter vertical line shows average location of voicing onset. Average audio signal amplitude indicated by gray shaded region. Right – electrode locations for clusters on the left.

The first Anatomically Constrained Phonological-to-Motor (AC PtM) processing cluster discussed is shown in green in Figure 15 (a). Activity ramps up quickly after stimulus onset and initially has a brief faster decay for 100 ms after the peak. After this the decay pattern slows down to a more gradual decay and has higher variance, returning to baseline levels after voicing completes. This AC PtM cluster is the most anatomically widespread, covering frontal, temporal, and parietal lobes in the left hemisphere. This results from the strength in the temporal similarity between the electrodes which is able to overcome the spatial distance measure which discourages this behavior. We name this cluster AC PtM diffuse ramp (AC PtM-dr). While activity is anatomically widespread, the majority of the electrodes fall in the frontal cortex. This area has been found to be heavily associated with language production and speech planning processes, including phonological working memory (Buchsbaum et al., 2005; Rottschy et al., 2012) in the left inferior frontal sulcus. The left premotor cortex is also seen to have a large number of electrodes present. This area has been associated with speech motor planning (Guenther et al., 2006; Guenther, 2016). Inferior somatosensory and posterior superior temporal

gyrus conclude areas outside of the frontal cortex included in this cluster, regions that have also been found to be active during motor planning (Flinker et al., 2015; Collard et al., 2016; Guenther, 2016).

The second cluster of the AC PtM group, shown in orange in Figure 15 (b), is the only one that displayed bilateral activations. This cluster is constrained to ventral regions of the brain in both the temporal and occipital lobes. We call this cluster AC PtM ventral ramp (AC PtM-vr). Both hemispheres had activity in the posterior interior temporal cortex, an area found to be associated with word and nonword readings (Vigneau et al., 2005). Active occipital regions have been described as visual word form areas, with activity in occipital and fusiform gyrus being found to correlate with reading comprehension and speech planning (Moore and Price, 1999). Activity in this cluster has the fastest activation and returns to baseline activity slightly quicker than the other clusters, which keeps it in line with visual reading comprehension.

The third cluster in the AC PtM grouping is shown in blue in Figure 15 (c). The activity pattern of this clusters is similar to AC PtM-dr and AC PtM-vr, but maintains a fairly constant and slower decay, as opposed to an initial quicker decay just after peak activation. This is seen as better fit as a linear decay as opposed to the slightly exponential decays of the previous clusters. Activity persists at a higher activity level than the previous cluster and lasts longer. Further, the cluster has a slight delay in activation compared to the other clusters in this group. Electrodes active in this cluster are localized to inferior regions boarding the frontal and parietal lobes, similar to the first cluster, but with more inferior coverage. This region includes the sensorimotor cortical

areas, which are well known and traditionally associated with motor planning and motor execution (Guenther, 2016). A single electrode is also seen in the posterior middle temporal gyrus. Since the bulk of the electrodes are in or near sensorimotor cortical areas we name this cluster AC PtM sensorimotor ramp (AC PtM-sr).

The final AC PtM cluster is shown in red in Figure 15 (d). This cluster is the most dissimilar from the other clusters. Activity in this cluster initially displays a symmetric pattern, ramping up just after stimulus presentation and decaying at the same rate, before exhibiting a slowly decaying and highly variable tail. We name this final cluster of the group AC PtM symmetric (AC PtM-s) due to this initial shape. The tail of this cluster is more similar with the ramp pattern and similar to the rest of the clusters in this group. This cluster has only a few electrodes in it, with all electrodes in anterior regions of the frontal lobe, an area active with reading comprehension (Buchweitz et al., 2009). This may help to explain the short duration and primary symmetric shape that ends well before voicing starts. The tail activity may be involved in phonological working memory, similar to electrodes in the first cluster that are in the vicinity of the inferior frontal sulcus (Buchsbaum et al., 2005; Rottschy et al., 2012).

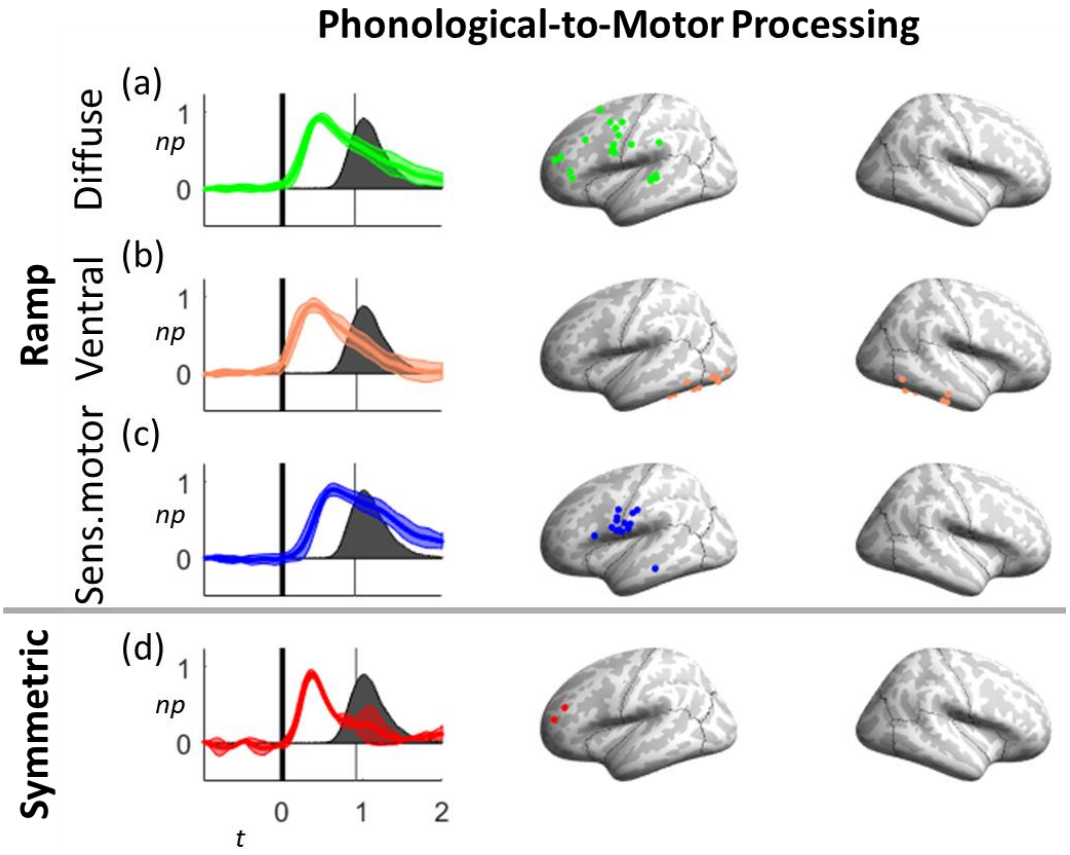


**Figure 15: Anatomically Constrained Phonological-to-Motor Individual Clusters**

Anatomically Constrained Phonological-to-Motor (AC PtM) results are broken into two characteristic time courses: (a-c) ramp and (d) symmetric. The ramp clusters are subdivided to three clusters: (a) diffuse, (b) ventral, and (c) sensorimotor. High gamma cluster activity patterns are shown on the left (normalized power (np) over time (t), in seconds), electrode locations for clusters on the right. Solid vertical lines indicate stimulus presentation and fainter vertical lines show average location of voicing onset. Average audio signal amplitude indicated by gray shaded region.

### *IV.3.iii Motor Execution*

The next grouping also is made up of four clusters, but all clusters in this group have bilateral activity. Further, all clusters are active during speech production, with activity starting prior to voicing and ending as voicing concludes, hence the group will be called Anatomically Constrained Motor Execution (AC ME). Results for this group are shown in Figure 16. Two of the clusters in this group, green and orange, are separated by

anatomical location with slight differences in the temporal activity duration, or broadness. The other two clusters show more distinctive shapes. The blue cluster shows a modification to the ramp shape, while the red cluster shows an initial period of suppression, both shapes that were not seen in CHAPTER III. A detailed description of each cluster is now provided, with the individual cluster results shown in Figure 16.

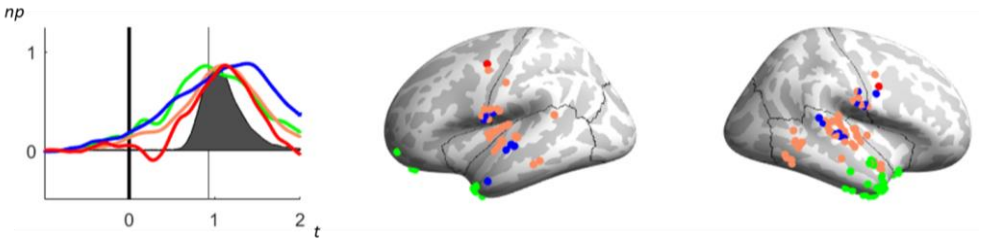


**Figure 16: Anatomically Constrained Motor Execution Clusters**

Left – Characteristic high gamma power (normalized power (np)) time courses (time (t) in seconds). Solid vertical line indicates stimulus presentation and fainter vertical line shows average location of voicing onset. Average audio signal amplitude indicated by gray shaded region. Right – electrode locations for clusters on the left.

The first AC ME cluster discussed is the green cluster in Figure 17 (a). The temporal profile of this cluster displays wide activation, with activity increasing above baseline immediately following stimuli presentation, peaking during voicing onset with a broad plateau, and then a symmetrical ramp down back to baseline levels after voicing has completed. We call this cluster AC ME symmetric broad (AC ME-sb) due to the broad activity duration. Anatomically, electrodes are located bilaterally within anterior temporal regions, with many electrodes falling on the ventral surface. Left hemisphere activity also shows some activity in anterior frontal regions.

The second cluster in the AC ME group is shown in orange in Figure 17 (b). This cluster shows a similar temporal activity pattern as AC ME-sb, but is more temporally constrained. This cluster is also centered over voicing and shows the symmetric pattern.

We name this cluster AC ME symmetric narrow (AC ME-sn). Electrodes in this cluster are located bilaterally within primary sensorimotor and auditory cortical areas. Electrodes surrounding the central sulcus are where motor and somatosensory representations of speech articulators are located (Penfield and Roberts, 1959; Guenther, 2016). Electrodes present in superior temporal cortex and inferior frontal cortex are located in regions found to be associated with auditory feedback for speech control (Guenther et al., 2006; Guenther, 2016) Additional activity is found in depth electrodes near Heschl's gyrus and planum polare, which have also been found active for speech production (Schönwiesner and Zatorre, 2009).

The third AC ME cluster is in blue in Figure 17 (c). This cluster shows an activity pattern that is a modification of the ramp shape, where slowly increases after stimulus presentation and quickly decreases after voicing. This is the inverse of the ramp activity pattern we has seen thus far. Hence, we name this cluster AC ME inverse ramp (AC ME-ir). Electrodes in this cluster are also located bilaterally within primary sensorimotor and auditory cortical areas, similar to AC ME-sn. Activity lasts longer in this cluster, with activity near the end being more consistent with the Anatomically Constrained Auditory Processing group, §IV.3.iv. The activity from this cluster may be involved in auditory feedback elements of speech production (Guenther, 2016).

The final cluster in the AC ME group, red in Figure 17 (d), is only seen in two electrodes. These electrodes are constrained to the anterior side of the central sulcus, with one electrode on each hemisphere. The shape of this cluster is very similar to AC ME-sn, as seen in Figure 16. The primary difference between the two is the brief dip in

high gamma activity 200 ms after stimulus presentation. This brief suppression in activity is another novel activity element not seen in any previous multi-subject clusters. We therefore name this cluster AC ME suppressed narrow symmetric (AC ME-sns). This cluster gets combined with the AC ME-sn without the anatomical constraint, but since the AC ME-sn center of anatomical mass is more inferior then this cluster these two electrodes get broken out into their own cluster.

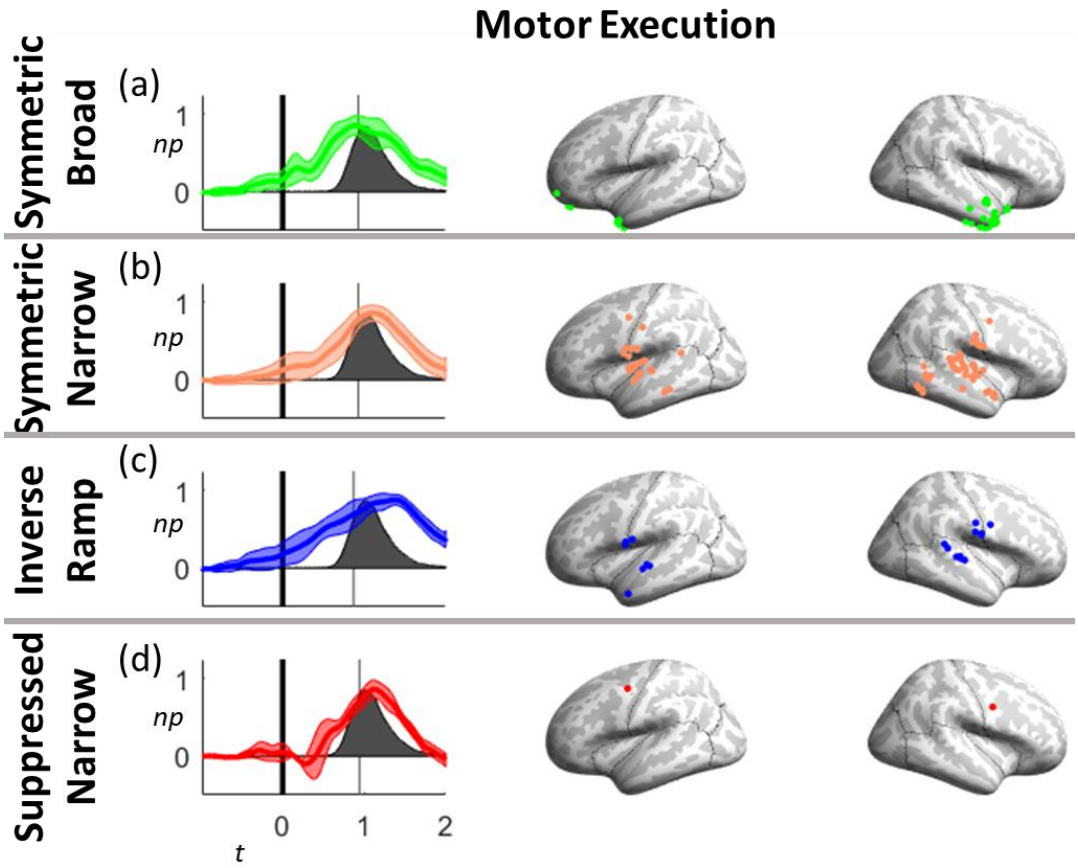


**Figure 17: Anatomically Constrained Motor Execution Individual Clusters**

Anatomically Constrained Motor Execution (AC ME) results are broken into two characteristic time courses: (a,b,d) symmetric and (c) inverse ramp. The symmetric shapes are subdivided to three clusters: (a) broad, (b) narrow, and (d) suppressed narrow. High gamma cluster activity patterns are shown on the left (normalized power (np) over time (t), in seconds), electrode locations for clusters on the right. Solid vertical lines indicate stimulus presentation and fainter vertical lines show average location of voicing onset. Average audio signal amplitude indicated by gray shaded region

### *IV.3.iv Auditory Processing*

The final Anatomically Constrained (AC) group consists of two clusters that fit the auditory processing stage. Hence we call this group AC Auditory Processing (AC AP). These clusters show activity that initializes with voicing onset. Activity peaks during voicing and ramps down after voicing has completed. Activity in these clusters differ primarily in their duration, with the green cluster lasting longer than the orange, as shown in Figure 18. Both clusters are primarily present in the auditory cortex. The orange cluster demonstrates bilateral activity, while the green cluster is limited to the left hemisphere. We will now detail the individual clusters, as summarized in Figure 19.

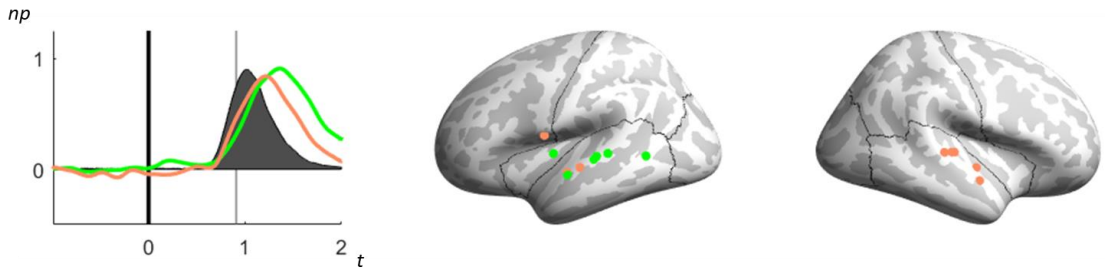


**Figure 18: Anatomically Constrained Auditory Processing Clusters**

Left – Characteristic high gamma power (normalized power (np)) time courses (time (t) in seconds). Solid vertical line indicates stimulus presentation and fainter vertical line shows average location of other alignment condition. Average audio signal amplitude indicated by gray shaded region. Right – electrode locations for clusters on the left.

Activity in the green AC AP cluster, Figure 19 (a), is only seen in the left hemisphere. The orange cluster, Figure 19 (b), has activity that is bilaterally seen in the temporal cortex. Activity in both is primarily localized to the superior temporal gyrus and bordering areas. Activity starts around the time of voicing onset and peaks just after the acoustic envelope peak. Both clusters have symmetric activity patterns. The orange

cluster peaks slightly before the green, and hence we name the orange clusters AC AP symmetric early (AC AP-se) and the green cluster AC AP symmetric late (AC AP-sl).

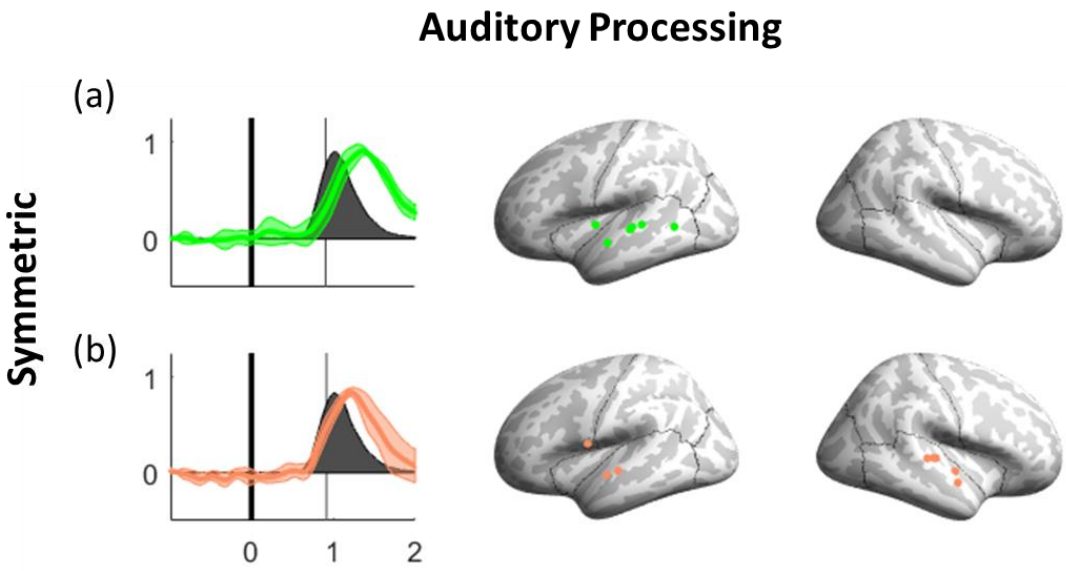


**Figure 19: Anatomically Constrained Auditory Processing Individual Clusters**

Anatomically Constrained Auditory Processing (AC AP) results come from a single characteristic time courses: (a,b) symmetric. The clusters primarily differ in their duration and peak locations, with (b) occurring earlier and (a) occurring later. High gamma cluster activity patterns are shown on the left (normalized power (np) over time (t), in seconds), electrode locations for clusters on the right. Solid vertical lines indicate stimulus presentation and fainter vertical lines show average location of voicing onset. Average audio signal amplitude indicated by gray shaded region.

### *IV.3.v Comparison to Unconstrained Findings*

In this subsection we will briefly compare the results of the anatomically constrained study back to those from CHAPTER III. Results from this study will be named Anatomically Constrained (AC) to distinguish them, and their temporal profiles will be shown in solid lines. We will refer to the results from CHAPTER III as anatomically-free and without the AC prefix, with the temporal profiles shown with dashed lines. This will be followed by a general discussion of the differences. This chapter only presents results for the stimulus aligned case, while the previous chapter

looked at both stimulus and voice alignments. For this comparison, we will focus only on the stimulus aligned case.

The Early Stimulus Processing (ESP) group is present in both analyses, each presenting the symmetric and ramp canonical temporal profiles, Figure 20 (a) with dashed lines for the anatomically-free analysis and solid for AC analysis. The symmetric activity patterns, green (AC ESP-s) and dark blue (ESP-s), align fairly well in the first half second after stimulus presentation. The ramp activity patterns, orange (AC ESP-r) and blue (ESP-r), also align well. Interestingly, the tails of the time courses align better with the opposite shape, e.g., ESP-r (dashed blue) has raised activity for the duration of the task similar to AC ESP-s (green), while both ESP-s (dashed dark blue) and AC ESP-r (orange) do not have long tails. There are some slight differences in electrodes, primarily due to the AC analyses, Figure 20 (c), only having activity present in the temporal lobe, while the anatomically-free case has some frontal activity as well, Figure 20 (b). Since this group consists of a limited number of electrodes, it is hard to draw hard conclusions about the comparison without more data.

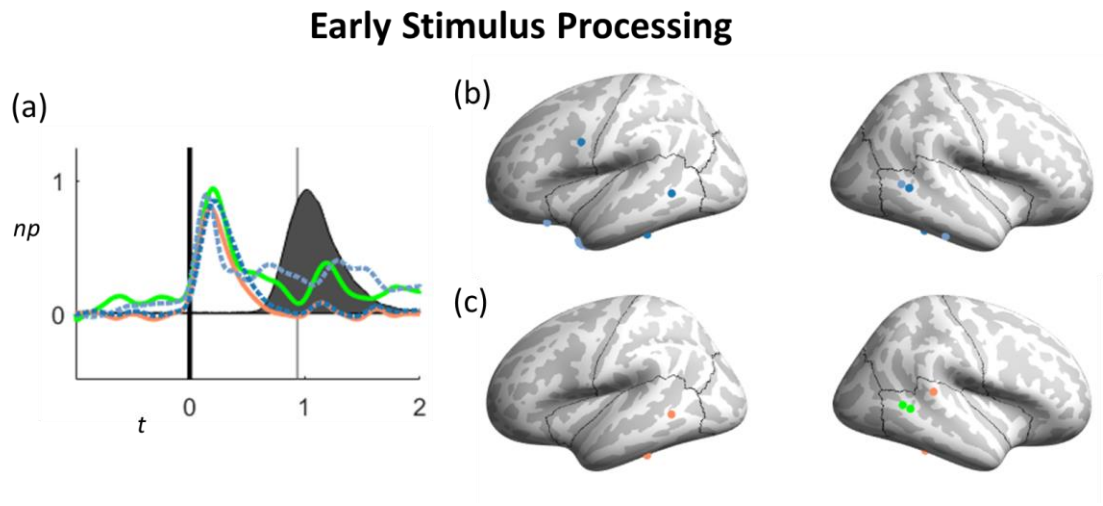


**Figure 20: Early Stimulus Processing Cluster Comparison**

Comparison of Anatomically Constrained (AC) and anatomically-free analyses shown with solid lines and dashed lines, respectively. (a) Time courses (normalized power, *np*, over time, *t*) from both analyses plotted together, retaining colors from their respective analyses – green (AC ESP-s), orange (AC ESP-r), dark blue (ESP-s), blue (ESP-r). (b) Electrode locations from anatomically-free analysis. (c) Electrode locations from anatomically constrained analysis.

In the phonological-to-motor (PtM) processing group, only a single cluster was found for the anatomically-free analysis for the stimulus aligned case. This cluster, PtM-r, has a ramp activity pattern, shown in dashed dark green in Figure 21 (a). The AC analysis had three ramp cluster. PtM-r is most closely related to AC PtM-dr (solid green) and AC PtM-vr (orange), as can be seen by the PtM-r having its temporal profile fit between these two, Figure 21 (a), and overlap in electrode coverage by comparing Figure 21 (b) and (c). There is limited overlap between PtM-r and AC PtM-sr, which has a more delayed activation and slower decay. The anatomical constraint in the AC analysis appears to separated AC PtM-sr from the Motor Execution (ME) group. ME grouping fits with the duration of AC PtM-sr activation, but previously a ramp shape was not found

within ME. Due to its activity pattern similarity with the PtM group it has been included here, but the temporal extent and active regions do hint at this cluster having some motor production functionality. The symmetric shape was only seen in the AC analysis, AC PtM-s (red), and only in two anterior prefrontal cortex electrodes. In the anatomically-free analysis a PtM symmetric (PtM-s) activity pattern was seen in electrodes also in the anterior prefrontal cortex.

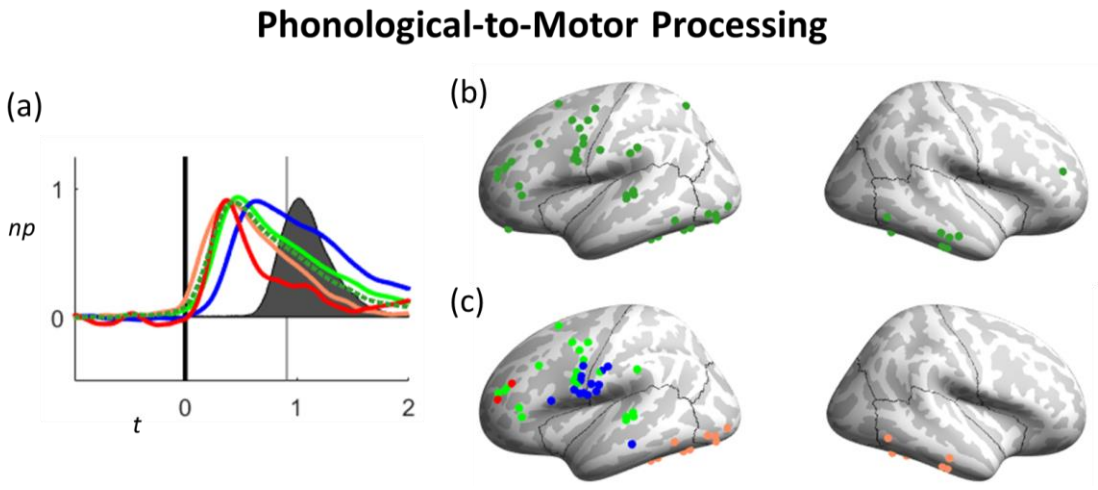


**Figure 21: Phonological-to-Motor Processing Cluster Comparison**

Comparison of Anatomically Constrained (AC) and anatomically-free analyses shown with solid lines and dashed lines, respectively. (a) Time courses (normalized power, *np*, over time, *t*) from both analyses plotted together, retaining colors from their respective analyses – dark green (PtM-r), green (AC PtM-dr), orange (AC PtM-vr), blue (AC PtM-sr), red (AC PtM-s). (b) Electrode locations from anatomically-free analysis. (c) Electrode locations from anatomically constrained analysis.

Figure 22 shows motor execution (ME) comparisons. ME-sn (dashed red) subdivides into two AC clusters, one showing the symmetric narrow pattern, AC ME-sn (orange), and the other showing a suppression in power before this pattern, AC ME-sns (solid red). AC ME-sns aligns very closely with ME-sn after the suppression, Figure 22 (a). AC ME-sn also aligns well, but with slightly broader activity tails. Anatomically

AC ME-sn overlaps with ME-sn inferior regions surrounding the central sulcus and superior and posterior temporal cortex, Figure 22 (b) and (c). ME-sn has additional activity in right anterior prefrontal cortex not seen in the AC analyses. The temporal profiles of the broad activity patterns, ME-sb (dashed pink) and AC ME-sb (green), almost line up exactly, Figure 22 (a). Anatomically, ME-sb is spread out to cover multiple brain regions, while AC ME-sb is constrained to anterior and ventral temporal cortex with some left anterior prefrontal cortex activity. Some of the missing activity from the ME-sb cluster in the AC analysis can be accounted for by the AC PtM-sr cluster, as discussed in the preceding paragraph. Other parts of the ME-sb activity seem to have fallen into the AC ME-ir cluster (blue). This cluster is located in traditional motor execution regions along the central sulcus and superior temporal gyrus. Averaging AC ME-ir and AC PtM-sr together produces a broad symmetric shape which is similar to AC ME-sb. The separation in the AC analysis of AC ME-ir and AC PtM-sr hint at motor execution elements that also functionally overlap with planning (AC PtM-sr) and auditory and somatosensory feedback (AC ME-ir). These feedforward and feedback speech production command pathways match theory (Guenther, 2016).

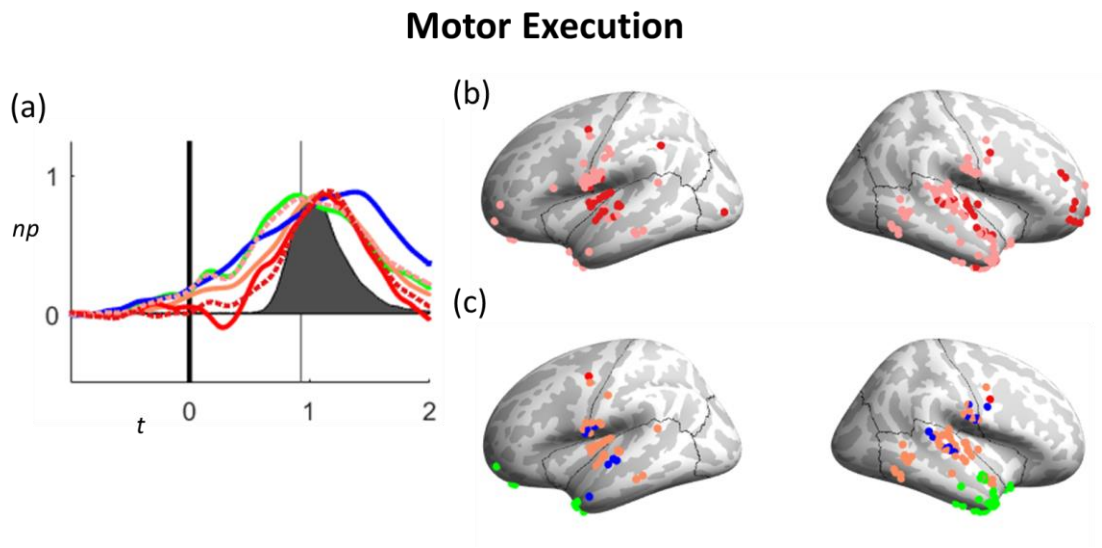


**Figure 22: Motor Execution Cluster Comparison**

Comparison of Anatomically Constrained (AC) and anatomically-free analyses shown with solid lines and dashed lines, respectively. (a) Time courses (normalized power, *np*, over time, *t*) from both analyses plotted together, retaining colors from their respective analyses – dashed red (ME-sn), dashed pink (ME-sb), green (AC ME-sb), orange (AC ME-sn), blue (AC ME-ir), solid red (AC ME-sns). (b) Electrode locations from anatomically-free analysis. (c) Electrode locations from anatomically constrained analysis.

Both analyses only had symmetric activity patterns for auditory processing in the stimulus aligned case. Activity is primarily located within the auditory cortex, Figure 23 (b) and (c). The two clusters in the AC analysis (AC AP-se, solid orange, and AC AP-sl, green) are primarily separated by temporal extent of the activity pattern, Figure 23 (a). AC AP-sl temporally and anatomically aligns closely with AP-s in the left hemisphere. AC AP-se shows a similar temporal profile as AP-s, but ends earlier. AC AP-se has bilateral activations, similar to AP-s.

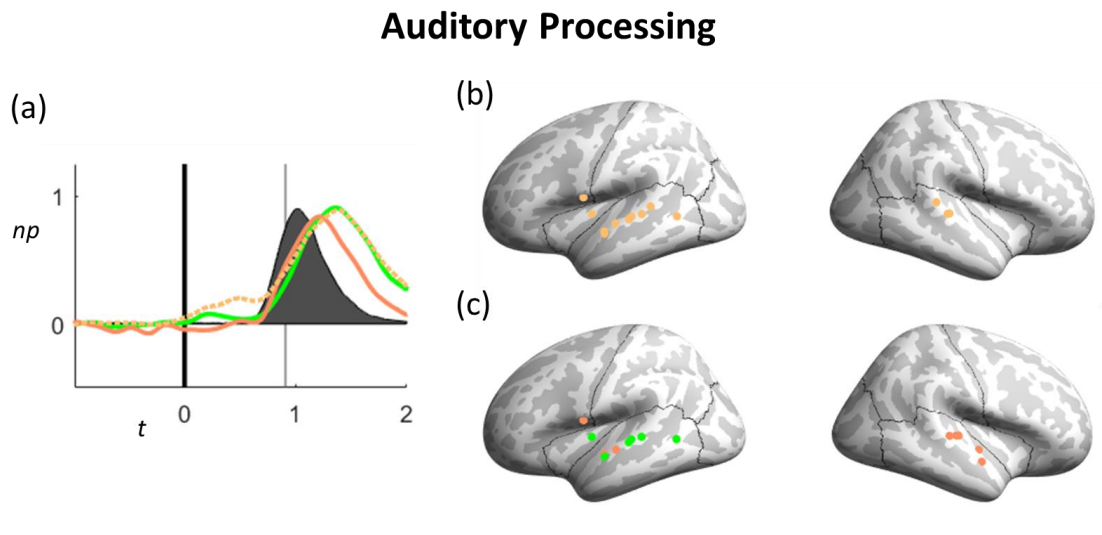


**Figure 23: Auditory Processing Cluster Comparison**

Comparison of Anatomically Constrained (AC) and anatomically-free analyses shown with solid lines and dashed lines, respectively. (a) Time courses (normalized power, *np*, over time, *t*) from both analyses plotted together, retaining colors from their respective analyses – dashed orange (AP-s), solid orange (AC AP-se), green (AC AP-sl). (b) Electrode locations from anatomically-free analysis. (c) Electrode locations from anatomically constrained analysis.

A couple observations are noted when comparing the results from the anatomically free and AC analyses. First, AC produces more clusters. This is due to stricter requirements placed on the clustering process that breaks up clusters that would otherwise be joined. Second, the additional component of the distance measure in the AC analyses also allows clusters to form with less similar time courses. This results in larger variances in the AC clusters compared to the anatomically-free clusters, especially in clusters with less than 10 electrodes. This also results in the generation of new characteristic time courses that are not in the anatomically-free analysis. Additionally, this causes some of the electrodes that were included in the anatomically-free analysis to not get clustered or get clustered in a different processing steps. Lastly, the AC analysis

sub-clusters reveal some slight timing differences between similar activity levels in different brain regions. Overall, there is a high degree of overlap in results between the two analyses, providing supporting evidence for the canonical activity patterns of CHAPTER III.

## IV.4 Conclusions

Overall, results from the analysis of this chapter follow those of CHAPTER III, which did not encourage spatial similarity between electrodes. The general canonical shapes of CHAPTER III, namely a symmetrical and ramp activity pattern, are also present within the constrained anatomical regions of this chapter. Further, the same four processing stages emerged from both analyses. There is a large degree of anatomical overlap between the two analyses as well, with the anatomical constrained analysis of this chapter largely being cortical subdivision of the anatomically-free analysis.

There were some differences in results in this analysis compared to those of CHAPTER III. Some new activity patterns were discovered and some of the electrodes did shift in what type of activity pattern, and in one case the processing step, they were involved in. This was primarily due to the smaller cluster sizes and electrodes per cluster in this chapter compared to the last, which also generated larger variances in the temporal profiles. Clusters were only restricted to consist of electrodes from multiple subjects, but did not have a restriction on the number of electrodes within a cluster. Thus, several clusters had very few electrodes in them. A larger subject sample size is needed to build enough evidence in the individual clusters and reduce the temporal variance. This would allow for better support of the results found in this chapter and check if they still hold.

The clusters that were most similar to those of the previous chapter had large number of electrodes, giving support to those findings, but additional research will need to be conducted to see if the more unique findings truly exist, like the inverse ramp and suppressed narrow activity patterns in the motor execution group.

This chapter presented results only for the stimulus presentation alignment case. This was done to show the validity of the method and not to intentionally mask any results from the voicing onset alignment case. As just mentioned, this analysis allows for breaking down clusters into smaller units, constrained to more confined anatomical regions, and hence creates more clusters with fewer electrodes in them. Data from more subjects is needed to dive into this type of more detailed analysis. In general, the voicing onset aligned analysis under with the additional anatomical constraint of this chapter produced the same findings as the stimulus presentation aligned case, namely it largely supported the findings of CHAPTER III with the primary difference being cortical subdivisions of the canonical temporal profiles. Some deviations were also witnessed in this alignment condition, but were contained within clusters with limited number of electrodes.

We hope the work of this chapter motivates others to continue this work and explore the functional and anatomical break outs that result from this type of analysis during speech and non-speech tasks. The findings that we present here hint at some fine timing and activation differences that exist between similar functioning clusters when they are analyzed with anatomical constraints. This could potentially help to reveal new findings, such as the propagation of activity across different brain regions, such as

differences in “what” and “where” pathways (Ahveninen et al., 2006). However, others who have looked at anatomical constraints have also noted an increase in variability compared to when the constraints are not place on the analysis (Berezutskaya et al., 2017). Therefore, it is perhaps appropriate to allow functional representations to be diffuse if that is what the data tells us, as in CHAPTER III. We encourage others to continue looking at both types of analyses, ideally with large subject sample sizes, to determine the correct spatial resolution to capture the canonical set of speech temporal profiles.

# CHAPTER V: Future Directions

This dissertation explored a new way to analyze electrocorticographic (ECoG) data recorded during speech, providing new insights into how the brain processes speech during various stages. A set of high gamma power canonical activity patterns, which capture local neural activity, were found across four processing stages: early processing of an orthographic stimulus, motor planning, motor execution, and auditory feedback to self-generated speech. This was done without any prior knowledge built into the methods on the speech network. Two types of activity patterns emerged: one taking a symmetric shape, where activity increases and decreases at the same rate, and the other taking a ramp shape, where activity decays at a slower rate than it activates. Additional, constrained analysis that limited the amount of anatomical spread a grouping of activity patterns showed similar findings as the unconstrained analysis, supporting the set of canonical activity patterns, or time courses, found.

This resulting set of temporal profiles provide unprecedented detail regarding the nature and timing of neural computations underlying the translation of phonological information into motor and acoustic output. The identification of characteristic time courses of neuronal activity during movement planning and execution has been more well studied for functions outside of speech and in non-human primates. These findings have provided critical insights into the brain mechanisms underlying motor control. Due to the relative lack of electrical recordings from the human brain, little is known about the temporal profiles of neuronal populations involved in uniquely human acts such as speech. This work utilizes electrical recordings form the human brain to provide some

new insight into the neural temporal profiles present during speech. Further, the analysis tools developed for this work provide a powerful means for identifying and quantitatively characterizing the neural computations underlying human speech production, with the potential to apply to other cognitive and behavioral domains.

This work branched into new areas, with some simplifications taken to provide this initial set of findings with some limitations present. A limited set of subjects, five, were used in this body of work. All of the subjects completed a single type of speech task: reading aloud monosyllabic utterances. The methods implemented novel elements, making some simplifying assumptions to build credibility in the method, but potentially limiting additional insights that could be gained. Conservative measures were used for electrode inclusion. Trends were simplified to linear models, via a Kalman filter, providing new insights, but not fully characterizing the activity patterns. Future research can build on the work of this study to provide additional insights and understanding of human speech and beyond. In the remainder of this chapter we present a few future directions.

## V.1 Large Dataset

One of the limiting factors of electrophysiology research, including speech, is getting access to human subjects to conduct tests. Speech, and in general language, are unique to humans and therefore necessitate the use of human subjects and limit the insights that can be gained from animal models. Further, intracranial electrophysiology is limited to patients undergoing clinical procedures who volunteer for research studies,

with speech being one amongst many types of studies trying to be conducted on a small subject population. Those who do volunteer are undergoing stressful procedures and are conducting the research experiment within a constrained amount of time and amongst a set of clinical evaluations, which may limit the quality of the data. Few alternatives exist, however, to capture the temporal resolution needed for detailed time analyses of speech without a new recording technique being invented.

In the current study five subjects were used. This was due to the small sample size of subjects who performed the speech task of the study. There were a total of six subjects who completed the task, but the seizure focal area of one of the was centered in speech critical brain regions and was therefore not used. This sample size is in line with other similar studies (Collard et al., 2016; Hullett et al., 2016; Leonard et al., 2019), but is too small to draw firm conclusions. It is our hope that others will continue this, or similar, research. With more subjects, stronger conclusions could be drawn. As time progresses it will be possible for meta-analyses to combine results across multiple studies to overcome the small subject sample sizes.

Along with a limited set of subjects, the current study also only utilized a single speech task, reading aloud monosyllabic utterances. This is another area for potential future work, to extend this type of study to other tasks involving speech production. There is a growing body of work in ECoG analysis of speech production and we hope that others venture into clustering-based research approaches such as that of this dissertation and (Leonard et al., 2019). We further hope others will use the methods developed within this work to look at other speech production and perception tasks, with

the potential to uncover new knowledge on neural time courses that exist under different conditions. For example, (Hamilton et al., 2018) found onset activated and sustained activity clusters present during continuous speech. The work of this dissertation may allow for additional insight into the temporal breakdown during continuous speech along with providing a means to quantitatively characterize it. Other recording methods, such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), have provided many insights into speech production through the collection of diverse experiments. We believe ECoG is on the brink of providing something similar, but is limited by the availability of subjects so will take some time. However, it is not a recording holy grail, as it has its own list of unique limitations, such as not recording in the sulci, inability to record or identify single neurons, and its use only in presurgical settings with limited and subject-specific coverage. Thus, the cycle of scientific discovery needs to continue to explore and develop new and emerging recording methods as well.

## V.2 Additional Analysis Focal Areas

The research of this dissertation focused on the discovery of characteristic time courses from ECoG high gamma power during speech production of orthographically presented monosyllabic utterances. There are numerous future directions that this line of research could go. This includes expanding to other frequency bands, such as using the power from the beta frequency band, with initial beta activity pattern results shown in APPENDIX D. Other frequency bands and time series representations have been shown

to have functional components to speech, motor control, and human cognition in general. The methodology of this dissertation opens up opportunities to apply similar techniques to reveal time courses present in other data representations.

CHAPTER IV begins to add some constraints to the analysis of time courses, by encouraging individual groupings to be cortically co-localized. Numerous prior studies, both within speech and in a broader context, have found local brain areas to process a task similarly, giving credence to adding anatomical constraints. Previous research has also found more variability when anatomical constraints are added versus not (Berezutskaya et al., 2017). The limitations of the number of subjects, as covered in §V.1, prohibits the resolution of the anatomical analysis from getting too fine, but this is an area that future research could further explore. As noted in CHAPTER IV, a simple radial basis function kernel was used to encourage locality within the clusters. Future work could focus on the more cortical-aware spatial kernels to exploit cortical representations such as anatomical differences in cellular structure or underlying cellular orientation to ECoG electrode placements. Hard limits, such as regions of interest, could also be enforced, but this approach would limit the ability to extract activity patterns that are naturally present as opposed to artificially created due to a boundary that may or may not line up with the underlying neural processing.

Finally, future research directions should branch out into other ways of looking at the data and task. This study only looked at trial data aligned by the presentation of the orthographic stimuli and aligning to voicing onset. There are many other ways that a speech production task can, and should, be broken down. This includes by analyzing the

type of stimuli presented and spoken (i.e., by phoneme, etc.), the location of articulation, and the familiarity or novelty of the stimuli, amongst many others. This all relies on having a large enough dataset to break the data up into these smaller units with enough trials supporting them.

### V.3 Speech Modeling

Intracranial electrophysiology, and electrocorticography in particular, is not a class of recording techniques that are readily used for speech production research. They are limited in their applicability, being used only with subjects undergoing clinical procedures who to volunteer to participate. It is therefore extremely beneficial to have alternative means to test theories and uncover new insights when fine timing details are desired. This in turn will help to guide and better use the limited ECoG subject studies. In the absence of animal models, speech neuroscientists have turned to models to provide this ability, such as the directions into velocities of articulators (DIVA) model (Guenther, 2016).

Speech production models have come a long way, mirroring the developments within the field. As new insights and knowledge are gained through human subject experiments, models are updated to reflect the findings and account for any differences that may exist. Now that we have more detailed time courses of speech production, a next step is to incorporate these findings back into our speech models, either by validating that the same results can be constructed from the model or identifying how to update the model to account for any discrepancy. Clustering analyses and detailed

temporal studies can then commence with data collected from models to explore different research avenues. This can then guide the development of future research paradigms to run with human subjects.

### V.4 Moving Beyond Speech

The focus of the current work was on speech production measured with ECoG. The methods developed for this research had that focus in mind, however are not inherently limited to applications involving speech, or even requiring ECoG data for that matter. Other studies that utilize ECoG could benefit from the analyses developed for this work, with the methodological similarity amongst studies related to other motor modalities. In particular, the implementation of an exponential distance measure for the temporal clustering analysis and the Kalman filter change point detection have potential to make strides in other ECoG applications. Moving beyond ECoG, the methods could be applied to other recording methods that have time series nature to them as well, with some rework on parameterizing the methods for the tolerances of the recording method. Moving beyond neuroscience, there also exists wider applicability as well.

# APPENDIX A: Kalman Filter

In this appendix, more details will be provided about the Kalman filter (Kalman and Bucy, 1961) and how it was adapted for the purposes of analyzing electrocorticographic (ECoG) data. The Kalman filter is a powerful tool for tracking time-series data and is typically used for prediction. It is not new to neuroscience, or even to ECoG research. It has been used extensively in ECoG BCI applications, including motor applications to predict movement trajectories (Gunduz et al., 2009; Li et al., 2009; Eliseyev and Aksenova, 2016) and as a means to provide task related frequency estimates (Gruenwald et al., 2017). Outside of BCI, its use is more rare. It has been used in limiting settings to help with modeling and prediction of ECoG signals, such as being paired with adaptive autoregressive modeling to predict seizure onset (Lie and van Mierlo, 2017). In speech modeling outside of ECoG, Kalman filters have been used to perform prediction and smooth parameter estimates over time to provide continuous synthesizer control (Guenther et al., 2009). In more general speech processing, Kalman filters have been used extensively for speech enhancement in the presence of noise (Grancharov et al., 2005; Mathe et al., 2012; Xia and Wang, 2015).

The Kalman filter is used heavily in statistics and control problems. It is based on the notion of tracking states of system by using a combination of observations and prior knowledge of the system, while accounting for several sources of noise, or uncertainty, in the model. The filter process is broken up into two primary steps, a prediction and update step. The prediction step uses a model to predict what the next state of the system will be. Then in the update step, this prediction is compared to the observation from the

system. Noise sources within the model include the uncertainty in the measurement of the observed state, the measurement variance, and model noise, which accounts for unknown or un-modeled dynamics of the system. The Kalman filter is an iterative algorithm that performs the prediction and update steps at each iteration to estimate the state of the system. Many references are available to more fully detail this algorithm that was first introduced in the early 1960s (Kalman and Bucy, 1961). In our application, we apply a novel application of the Kalman filter to ECoG time series data, as described in CHAPTER III (§III.3.ix). In particular, we use Kalman filters to track the time course of ECoG high gamma power during a speech task.

In our application, the Kalman filter is initialized during the trial baseline activity, which is the non-speech period prior to each trial. Thus, the initialized Kalman filter is set to represent neural activity during non-speech functionality. The filter parameters are then adaptive updated as time goes on, one of our novel contributions. This is done using a learning rate, which relies more on the model estimates, i.e. the prediction, and less on the observation, i.e. the update, as time progresses. The inclusion of the learning rate enables the filter to detect changes in activity trends, as opposed to trying to track through them. It has been shown that reduced variability in ECoG signals is present during stimulus onsets (Dichter et al., 2016), and thus we apply the learning rate to the Kalman filter gain, locking in both the activity trend and model variance to detect changes. Our rationale for this change is that as the model more accurately captures baseline activity we can rely more on the empirical data of the past to predict the future.

Unlike most applications of Kalman filters, we are not interested in creating a track that is capable of modeling the entire time-series representation, but instead we wish to model the several discrete trends that appear throughout the duration of the signal. This requires two additions to the Kalman filter model: 1) the ability to detect when a trend has changed and 2) the ability to start a new Kalman filter at the point of change to model the new trend. In the preceding paragraph, we described how we modify the Kalman filter with a learning rate. The learning rate allows us to accomplish our first addition by putting more emphasis in the prediction step as time goes on, which allows the estimated signal to deviate from the observed signal as the trend changes. Next, a threshold is added to detect that the difference between the prediction and observation is too large, resulting in a change. After a change is detected, a new Kalman filter is started, consisting of our second addition to the Kalman filter model. The new filter is started with a modified version of the initialization procedure, as described in §III.3.ix. We provide some additional details and rationale in this appendix.

Several others have looked at Kalman filter change detectors with varying degrees of similarity to our approach. In (Lee and Roberts, 2008), extreme value theory is used to set the change criteria for the Kalman filter. Lee and Roberts track the run length of the filter and apply a one-sided test to determine change. Results showed promise, but the method fails to detect changes that are closely spaced or when the trend returns to a baseline period. In (Soule et al., 2005), a Kalman filter is first used to capture normal traffic for an internet service provider (ISP) traffic pattern. Anomaly detection is then performed on the residual of this prediction to identify changes. In (Severo and Gama,

2006), a Kalman filter is used as a regression predictor. The residuals of the Kalman filter prediction are compared to the observation are evaluated to determine if a change occurred, forming a change detection from the cumulative sum of recursive residuals. Finally, in (Moussakhani et al., 2014), a Bayesian framework is utilized to place prior knowledge on the expected nature of the change to improve change detection performance. With this Bayesian criteria, Moussakhani et al. showed that their detector becomes a matched filter when the uncertainty in the change goes to zero and becomes an energy detector in opposite case, when the uncertainty in the change goes to infinity.

Our approach uses the residuals between the prediction and observation to detect change, similar to previous methods, but instead of specifying the bounds of the change detector or formulation of it ahead of time, we employ a learning rate that adaptively changes the bounds of the detector as more observations are used to update the filter. This slowly adapts the change detector over time to capture the trend of previously seen observations and evaluate how well new observations fit that trend. Learning rates are not novel, and have been used on ECoG data in the past in conjunction with a Kalman filter (Hsieh and Shanechi, 2018). We apply a learning rate only to the Kalman gain factor, for the purposes we just described, and do so in the context of a change detector. This approach is novel, but relates to the previous methods just mentioned.

The rest of this appendix is broken into two sections. In §A.1 we provide more details on the methodology developed for the Kalman filter change point detection. In §A.2 we discuss some of the alternative methods that were considered.

## A.1 More Details on Methodology

### *A.1.i Illustrative Example of Methodology*

We begin first by providing an illustrative example to provide better understanding and intuition on our approach, refer to Figure 24. The high gamma power temporal profile from one electrode is shown on the left side of the figure, with stimulus presentation at zero on the x-axis. The right side shows how the Kalman filter method segments this time series into a discrete set of trends. A yellow bounding box has been placed around the baseline, non-task period, 1 second to 500 milliseconds prior to stimulus presentation. This is the activity that is used to initialize the Kalman filter. The Kalman filter (red) can be seen as having a flat trend during this period. It is just above zero, and not at zero, due to rescaling of the signal after z-scoring (§III.3.vii).

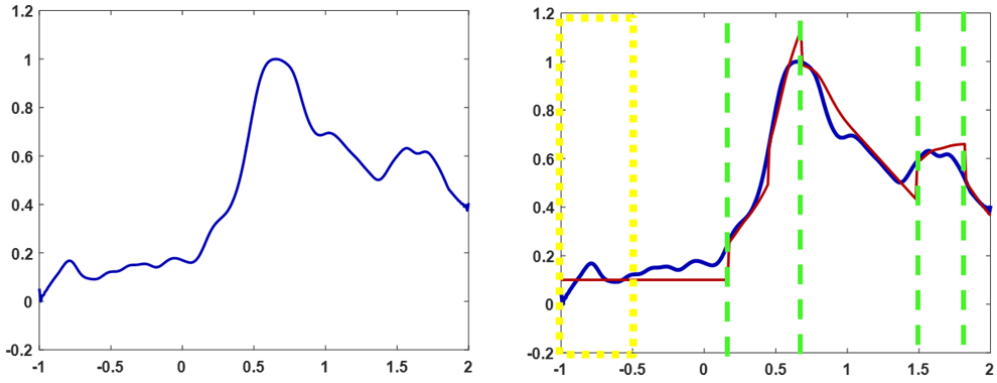


**Figure 24: Example Kalman-Filter Segmentation**

Left: High gamma signal to be segmented into discrete trends. Right: Kalman filter output of trend segments. Yellow box denotes the baseline period that is used to initialize the filter. Red line shows Kalman filter estimate. Vertical green lines mark change point locations. x-axis: time (seconds), y-axis: normalized power

The learning rate limits the amount of new information, observations, that are used as the filter progresses past the baseline period. This has the effect of maintaining

the learned trend from the baseline period, as seen by the flatness of the filter (red) which does not track the high gamma activity (blue) as it starts to increase. A green vertical dashed line indicating a change point is seen around 100 ms after stimulus presentation (x = 0). At this point, the new activity observations (blue) falls outside the tolerable bounds of the model estimate (red). A change in trend is therefore declared and a new Kalman filter is initialized. This is seen by the discrete jump in the estimate (red), which has different trends before and after the change (flat before and positively sloped after).

As time progresses the new estimate adapts to represent the underlying observations and slowly the learning rate locks in this observed trend and prevents the filter from changing too much. The filter adapts more to the update step earlier in the trend, i.e. the nonlinear changes in the estimate (red) around 450 ms, and less so as time progresses, i.e. after 500 ms. This new trend is found to no longer fit the data at a time of around 650 ms and another change is detected. A new filter is initialized and this process continues for the duration of the trial.

Four changes are detected in this example, resulting in 5 discrete trends. The first trend is flat representing pre-task baseline activity. Just after stimulus presentation activity is positively sloped up as more task-related neural activity starts. This peaks around 650ms and is followed by a downward trend as activity start to turn off. A new trend is seen from 1.5 seconds to 1.8 seconds after stimulus presentation, potentially due to auditory feedback to self-generated speech briefly increasing activity levels. After this, activity again decreases back to baseline levels as the final trend. Quantitative values for the trend in each segment are obtained from the Kalman filters.

This illustrative example helps to explain the rationale of the approach and the strength that it provides in aiding in characterizing activity patterns. It also illustrates the ability to identifying the discrete points of trend change and quantitatively model each trend segment. The method provides the potential for additional insights into neural mechanics beyond more traditional static statistical significance analysis that are commonly used. We discuss some of the algorithm elements that warrant a more discussion.

### *A.1.ii Kalman Filter Model Choice*

One of the big design questions is what form the Kalman filter should take. This is heavily dependent on the domain and what it is that is trying to be predicted. In our application, we are trying to characterize the trends and discrete changes that occur as neural activity respond to a task, as measured from ECoG high gamma power.

High gamma power is active during tasks and relatively inactive during resting state, refer to §II.2. Our task baseline period can be thought of as a resting state and therefore should have a trend that shows a lack of activity. A first approach to the Kalman filter is to have the filter track a constant value that captures this baseline period. While this would work for the baseline period, it will not provide any insights into what is happening when activity is different than baseline. This is a shortfall of earlier methods, which treated all of the analysis as a comparison to baseline, which can be thought of as a filter tracking a constant value, refer to CHAPTER II. This does not allow for assessing or characterizing non-constant activity, i.e. activity ramps.

The natural next level of complexity to add is to allow for tracking of a linear representations, i.e. a bias and a rate of change. This is the form of the Kalman filter that was used in this work. Going beyond this, to quadratics, polynomials, or other nonlinear representations may have helped to better characterize some of the individual trends, but suffer from a loss comparison across trends and, importantly, interpretability. Hence, after exploring some different functional representations we settled on having the Kalman filter perform linear modeling of trends, as provided in Equation 2 in CHAPTER III. This sets the basic form of the Kalman filter, as detailed in §III.3.ix.

### *A.1.iii Learning Rate & Change Detection*

A learning rate was used to slowly lock in learned trends and variances in the temporal profile and prevent adapting too much to new observations. This also provides the ability to identify changes. Over time, the learning rate shifts the expression to weigh the model prediction more than the observation of the high gamma power.

The learning rate takes the form of a decay, allowing the power rate to be learned from the data during filter initialization and less so as it matures. The decay rate is set by single fixed parameter that was selected to be 100 ms. This parameter sets the decay to be on the order of phonemic expression, which is one of the base units of speech and is on the order of 100 ms (Stevens, 2000). The decay is applied to the Kalman filter gain factor, which sets how much the update step, or observations, contribute to the filter estimate. Pseudo-code for the algorithm is provided in Table 3, illustrating this modification to the Kalman filter. Refer to §III.3.ix for a more complete description of the terms.

**Table 3: Kalman Filter Pseudo-Code**

```
# Setup
threshold = compute_threshold()
A = [1 1; 0 1] # State transition matrix
H = [1, 0] # Only measuring power
decay_param = 100 ms
Fs = 1000 Hz

# Initialization
# X = [Power, power rate]
X(:,1) = [mean(power(training data)); 0]
Q = variance(training period sample differences)
R = variance(training data)
P = variance(power sample difference)
r = 0 # Empirical rate: Baseline period flat

For i = 1:length of trial

   # Kalman Filter Prediction
   X(:,i) = A*X(:,i-1)
   P = A*P*A'+Q

   # Change Detection Check
   if abs(Z(i) – X(1,i)) > threshold
      # Declare change
      re_initialize_Kalman_filter()
      # reset number of points in current trend
      num_points = 1

   # Kalman Filter Update
   r = compute_empirical_trend()
   # alpha = Learning rate (decay)
   alpha = exp(-num_points/(decay_param*Fs)
   K = P*H'*inv(H*P*H'+R)
   K = alpha*K
   X(:,i) = X(:,i) + K*(Z(i) – H*X(:,i))
   num_points = num_points + 1
```

The following sections describe some of the functions that are used, including change detection (§A.1.iv), starting a new filter after a change (§A.1.v), and characterizing the resulting trends (§A.1.vii).

### *A.1.iv Change Detection Threshold*

Setting the threshold is briefly covered in CHAPTER III (§III.3.ix). Here we describe it in a little more detail.

An inverse Q-function is used to compute the threshold (Borjesson and Sundberg, 1979; Beaulieu, 1989; Craig, 1991; Karagiannidis and Lioumpas, 2007; Simon, 2007). Since the noise sources are assumed to be Gaussian, the estimate of the filter state can be thought of as a Gaussian random variable. The Q-function uses this to construct a normal cumulative distribution function (CDF) for the Gaussian distribution of the estimate, or more accurately one minus this CDF. Putting it into this representation allows us to set a bound on the acceptable values that would be drawn from the same distribution. A threshold can also be set on the values at the tails of the distribution that do not fit the CDF, at which point we would declare a change in the underlying distribution that the data is drawn from. The data referred to in this case is the difference in the observed values of the high gamma power with the estimate of the filter.

We explored different bounds to use, but settled on using a 95% bound. This results in estimates that fall in the 2.5% tails of the distribution being determined to not fit the trend and a change declared. The inverse Q-function of 95% is taken and then scaled back from the normal distribution to the empirical distribution of the high gamma power to get a threshold value in the units of the observed activity. Scaling is done using the covariance estimate for high gamma power during the baseline period, which is also used in the Kalman filter as $R$. This gives us a confidence bound for activity within a trend for individual electrodes, as provided in Equation 9.

**Equation 9: Electrode Confidence Bound**

$$threshold = \sqrt{R}\, Q^{-1}\left(\frac{1 - bound}{2}\right)$$

The desire is to have a global threshold that can be used across all electrodes to help mitigate confounding factors such as multiple comparisons. Thus, a distribution of the confidence intervals over all electrodes is taken, resulting in a beta distribution. From here, a conservative confidence interval of 99% is taken to use as the global threshold. This makes up the step of computing the threshold in the pseudo-code of Table 3, *compute_threshold()*. The threshold functions to determine when there is a change in the underlying trend, using an empirically driven bound that was determined using baseline activity from all electrodes that are all scaled by their respective z-scores to have a common distribution.

### *A.1.v Re-Initialization*

As noted, when a change is detected a new Kalman filter needs to be reinitialized to track the new trend. This process is described in CHAPTER III (§III.3.ix) and captured in the pseudo-code of Table 3 as *re_initialize_Kalman_filter ()*. Section §III.3.ix describes using 10% of historical and 90% of future observations from the point of change to reinitialize a new filter. The duration of this is again set to be that of phonemic expression, 100 ms, with 10 ms of prior observations and 90 ms of future observations. The split between historical and future observations was experimentally found to be a good split, as the trends typically would start to show a pattern of change slightly before the change threshold was crossed and a change detected.

If there is not enough data to reinitialize the trend, it is approximated with the amount of data that is available within these bounds. This creates some undesired behavior at the edges of the trials, but changes at the edges are likely not due to the task and not analyzed, hence special edge cases were not required to be handled separately.

The empirical trend of the data is the rate of change of the observed power for the current trend under analysis. It is computed as the difference in the power from the current time point to the beginning of the trend divided by the elapsed number of time steps in the current trend. Note that the power itself is the only measurement that is directly observed. In the pseudo-code, Table 3: Kalman Filter Pseudo-Code, the empirical trend is calculated in *compute_empirical_trend()*, which is used in the update step for the power rate. During re-initialization of the Kalman filter, or beginning of a new trend, the rate is computed according to the description provided in the beginning of this subsection.

### *A.1.vi Change Points*

In our discussion in CHAPTER III (§III.3.ix), we describe the discrete changes as change points. We do so following prior work of similar approaches used in other domains (Page, 1963; Lavielle, 2005; Haynes et al., 2017). There is a class of algorithms called online change point detection that have provided a nice framework to detect when sequential data exhibits a change and are cast in a probabilistic framework (Adams and MacKay, 2007). We started with a modified version of this approach, as discussed later in §A.2. We migrated towards a Kalman filter as insights were gained. Kalman filters have been shown in the past to provide the ability to perform change detection (Lee and

Roberts, 2008), and hence we are not novel in the general concept. Instead our use is novel in our application and the domain, as well as the way in which we modify the Kalman filter to perform change detection as discussed in §A.1.iii.

### *A.1.vii Characterization*

The primary use of the Kalman filter was to characterize the time courses and provide a way to describe the task related activity patterns. Two recurring activity patterns were discovered with ramp and symmetrical shapes, as described in CHAPTER III. In CHAPTER IV some additional patterns were seen with the addition of a spatial cortical constraint. Individual subject clusters also showed some new patterns, in particular a high gamma power suppression shape, refer to APPENDIX D.

These patterns were characterized and named based on the findings from the Kalman filter analysis. To illustrate this, Figure 25 shows the ramp and symmetric shapes on the left and right, respectively. CHAPTER III discusses the criteria for naming cluster shape, but with the illustration provided here we aim to help with intuition. Notional trends are overlaid in red to visually show the differences in the rate of change. The left plot shows the ramp pattern. The rate of the ramp up (segment between the first two vertical orange dashed lines representing change points) is 2.8 times faster than the decay rate (segment between the next two change points). This illustrates the asymmetrical rates that characterize the ramp shape, where the activation is faster than the decay. The symmetric shape is shown on the right, where activity increase and decrease are about the same, with the increase rate only being 9% faster than the decay rate.

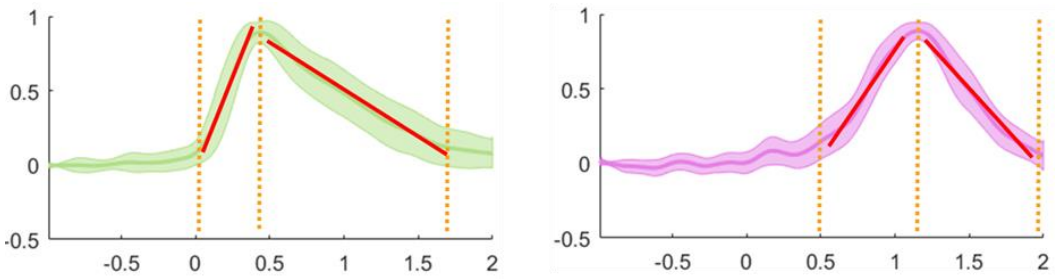


**Figure 25: Exemplar Shape Trends**

Detected changes indicated by vertical orange lines (i.e. change points). Red overlay lines illustrate power rate for activity increase and decay trends. Left: Ramp pattern (stimulus aligned phonological-motor processing cluster from CHAPTER III). Right: Symmetric pattern (stimulus aligned auditory processing cluster from CHAPTER III).

The Kalman filter change point detector provides a novel way to provide activity characterizations and has the potential to shed new insights into how neural populations are recruited for a task as well as return to baseline activity, and the differences between these two. Further, it allows for a comparative analysis between different activity patterns, as additional understanding may be found by looking at the differences in the trend rates. Some initial comparative analysis was provided in §III.6.

Lastly, a final comment on the shapes characterized in this study. We do not believe we are the first to see these shapes, but we do believe we are the first to describe them and come up with a formal way of doing so. In CHAPTER III we commented on how these shapes can be seen in previous research by a visual inspection of the figures. It can be seen in some of the prior work that the ramp and symmetrical shapes are present. They are, however, not often seen together in the same study due to the methodology used that smooths the temporal profiles and does not emphasize the significant portions of the signals, as discussed in more detail in APPENDIX B.

## A.2 Alternatives

In this section we touch upon alternatives that were considered for change detection and trend analysis.

Prior to developing the Kalman filter-based change detection, we initially approached the problem by modifying a Bayesian change point detection algorithm as described in (Adams and MacKay, 2007). This approach looks for abrupt variation in the generative parameters of a data sequence through its implementation as an online clustering algorithm. At each time point, the current observation is evaluated to see if it fits the cluster of the previous observation, an inactive cluster (i.e. from historical observations), or an altogether new cluster, with the last two representing change points.

The method is parametrized by several values, most important being the expected cluster noise variance. A cluster containment probability is computed to set the probability mass around a cluster, specifying acceptable ranges for data to fall within the cluster. Similar to the Kalman method used, we use a Gaussian assumption and parameterize similarly. Also similar to the Kalman method, an initial state needs to be setup from "training data". The baseline period was also used to initialize this method, with the design enforcing the baseline period to be clustered together.

This is an online method, so it is an iterative process similar to the Kalman method. At each time step, the observation is evaluated to find the closest cluster. If the closest cluster is the same cluster as the last observation, then the distance between the observation and the cluster is computed. The distance is a statistical measure using the cluster mean and variance under a Gaussian assumption. A threshold needs to be

specified ahead of time on what is an acceptable distance. If the observation is within this limit, it is added to the cluster. A "soft" change point is calculated based on cluster fit, thus providing a probabilistic representation similar to soft clustering algorithms. The cluster is updated with the observation.

If the distance exceeds the threshold or if the observation best fits with a previously seen cluster, a change point is declared. The observation is either is added to a previous cluster or a new cluster is initialized, depending on fit. For our purposes, we did not care about matching previous clusters beyond the currently active cluster. Thus, we implemented a forgetting factor to remove any previous clusters that are not active, creating the desired situation of being able to determine if the data point either fits within the current trend or if a new trend needs to be declared. The variance and threshold were set using the same logic in the Kalman method, refer to §A.1. An additional component was incorporated to prevent the ping-ponging between clusters. The addition maintained deactivated clusters in memory for 10 ms to see if the newly activate cluster should be merged with it.

The behavior of this approach displayed what was desired, but additions were needed to create better "smoothness" of the resulting cluster, i.e. ensuring similar trends were grouped together. The method was sensitive to steep rate changes, i.e., as seen in the ramp shape. However, it set the groundwork for the eventual Kalman filter method and helped to motivate going in that direction, with a lot of the elements of the Kalman filter drawn from the preceding work done with this method.

An example result from this method compared to the Kalman filter method is shown in Figure 26. The top plot displays the results of the Kalman filter, where it is seen that the change points (vertical green lines) line up fairly well with the underlying activity (blue curve). The bottom plot shows the Bayesian change point detection method, with change points seen lagging the actual activity trend change. In experiments, this could be fixed in specific instances by modifying the threshold criteria, but this resulted in other activity patterns having way too many change points declared. Thus, a tradeoff formed between too many changes and changes that lagged the actual change. The Kalman filter method produced far more stable results that fit the data and did not show wide variability from temporal profile to temporal profile.

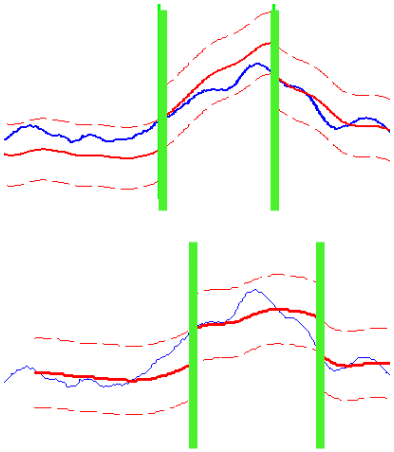

**Figure 26: Kalman Filter vs Bayesian Change Point Detection**

Same underlying temporal profile (blue) used in both plots. Predictions from the methods shown in solid red, with bounding regions for detecting changes in dashed red lines. Change points are marked with vertical green lines. Top: Kalman filter method. Bottom: Bayesian change point detection method.

Other model-based methods were attempted to fit the shapes of the activity patterns as a way to describe the trends. Methods explored in this category included fitting splines, polynomials, and sigmoid functions to the data. With an explicit

expression for the function, i.e. spline, it was thought that trends could be discussed in common terms, i.e. in the form of the function parameters. This would then create a common language for trend description. None of these methods produced satisfactory results without heavy additional preprocessing. The preprocessing in turn removed the fine timing information desired in this study.

Many variants of the Kalman filter have been developed over the years, including the extended Kalman filter (EKF) and unscented Kalman filter (UKF), which are nonlinear variants (refer (Julier and Uhlmann, 2004) for a review). We initially explored these options, however, we decided to stay with the base Kalman filter model for its descriptive ability. The work of this dissertation was a new foray into how to characterize and describe ECoG data, so we determined that maintaining linear representations provided the best initial framework for describing trends and moving beyond simple identification of statistically significant changes.

The Kalman filter enables describing trends as increasing / decreasing and quantifying their rate of change, with the underlying tracking logic set in the same measurement space that we are characterizing the trends in. Going beyond this would require more work to figure out how to describe the results, but would provide the ability to describe the nonlinear properties, i.e. transients, immediately around the change points. Going to nonlinear representations may allow for better insight into what is happening in plateaus and how trends transition, but getting to that point requires the initial research we have performed to set a first step in this direction. We hope our work here motivates others to continue this line of research.

# APPENDIX B: Clustering

In this appendix, more details are provided about the clustering methods used in this dissertation. The methods described herein were applied to electrophysiology recordings the surface and depth locations within the brain, but can be used more generally for time series clustering. This appendix does not describe any preprocessing steps or other elements of the methods outside of what is needed for the clustering. For a review of preprocessing steps, refer to CHAPTER III and APPENDIX C. The signals used for clustering are the preprocessed electrocorticographic (ECoG) high gamma power task-related temporal profiles that have been averaged across trial, i.e. event related spectral response (ERSP), z-scored using the non-speech task baseline, and down-sampled to 100 Hz. These steps are not necessary for this analysis, but are the core preprocessing steps used for most of the analysis in the chapters.

Clustering is a machine learning technique that segments and groups data, with the resulting groupings referred to as clusters. Further, clustering is a form of unsupervised learning. In unsupervised learning there are no labels for data points and instead the learning process identifies patterns and relationships in the data. In the case of clustering, the learning process involves identifying similarity in the data and grouping the data based on the similarity. Thus, the goal is to cluster the data in such a way that data points that fall within each cluster are similar and data points between clusters are different. The measure of similarity is a critical component of clustering and will be discussed in detail throughout this appendix.

We motivate our use of clustering for determining characteristic time courses of speech from ECoG due to the mechanisms of clustering, the properties of ECoG, and the underlying neural processes of the brain, as discussed later in this appendix and in CHAPTER II. The underlying neural dynamics of speech are not understood well enough to know what the characteristic time courses are, hence lending the analysis and discovery of time courses to an unsupervised approach. Taking a data-driven methodology to figure out the time courses is thus appropriate, as it does not make undue assumptions and instead lets the patterns and relationships naturally present within the data to drive the findings.

Clustering can be applied to various representations of the data. In this research we wanted to capture the underlying neural dynamics, so we chose to work with the data in as close to its raw form as possible. This means that we do not cluster over parametric models constructed from the data or from features extracted from the data, refer to §II.3.i for a discussion of studies that take this approach. Instead, we cluster directly on the preprocessed signals. We use the time domain representation of the signal, as opposed to the frequency domain, as we are interested in the fine temporal details and relationships between activity from different electrodes.

One important component of clustering is the measure that is used to assess the degree of similarity, or dissimilarity, between the data points. This typically comes in the form of a distance measure. For some applications the notion of what a distance should be used is fairly straightforward. Time series analysis brings extra difficulty to measuring distance, however, as the distance needs to be computed over vectors of

ordered values instead of scalar points. Thus, caution needs to be taken when working with time series representations, as there are interdependencies in the ordered list. These interdependencies may or may not be important to emphasize in a distance measure and selecting the wrong distance measure could result in suboptimal clustering performance that does not capture the underlying structure of the data that is intended to be grouped. Preserving time (temporal position, scale, and order) is an important factor in our research, as we aim to understand the underlying time course of speech so do not want to warp time. Our choice of distance measure will be discussed in §B.1.i.

Many different time series clustering approaches exist, refer to (Warren Liao, 2005; Aghabozorgi et al., 2015) for a more comprehensive review. We will now briefly situate our approach within this work, while providing rationale. We are interested in using the entire time series sequence (over the trial duration) in this work. Other time series clustering options include using subsequences, time points, or features of the time series to cluster, none of which suit the purpose of this research in getting characteristic time course over the entire duration and from a high gamma representation. Further, we are interested in shape-based clustering approaches, where we aim to find similarity in the shapes of the activity patterns from the different electrodes. We use a completely unsupervised approach to let the data determine what the activity patterns are and how they are clustered, since we do not have any strong priors on what characteristic activity patterns should look like. Hierarchical clustering approaches are attractive method to accomplish this data-driven discovery. A subclass of hierarchical clustering, agglomerative clustering, starts with every data point as their own cluster and then

systematically update clusters one at a time by iteratively merging clusters based on closest distance. This approach uses the pairwise distances between data points to cluster and hence fails to adapt to evolving notions of clusters as they are formed. Partitioning clustering approaches on the other hand form *k* groups by finding the best combination of the data into *k* clusters. This uses the full dataset to form *k* clusters, versus hierarchical clustering which would only use two of the *k*+1 clusters, to update cluster definitions to result in *k* clusters. Representative time series, or characteristic activity patterns in our terminology, result from the means of each cluster. We will turn to a hybrid approach that combines hierarchical and partitioning approaches, leveraging the iterative structure of hierarchical, but using more of the dataset to update cluster definitions. More details will follow in §B.1.ii.

One of the challenges with clustering analysis is determining how many clusters exist in the data. Agglomerative hierarchical clustering proceeds until all clusters are merged into a single cluster. This requires a way to assess performance to know which step of the hierarchical clustering to use to get the number of clusters and their characteristic temporal profiles. Partition clustering similarly suffers, requiring *k* to be defined. Furthermore, multiple versions of the clustering can be run from a specified value of *k*, necessitating the need for a way to determine selecting the best version. We address this challenge of selecting cluster size in §B.1.iii.

In the remainder of this appendix we will go through the specific details of our clustering approach. In §B.1.i, we discuss the novel distance measure we develop to highlight aspects ECoG data that are not covered in existing measures but crucial for

interpretation. We then discuss the hybrid iterative clustering approach we employ that uses elements from hierarchical and partitioning approaches in §B.1.ii. We follow this with details on how we select the number of clusters in §B.1.iii. After concluding our clustering approach in §B.1.iv, we provide a discussion on alternatives pursued and tested in §B.2.

## B.1 More Details on Methodology

### *B.1.i Distance Measure*

After surveying existing distance measures, it was determined that a custom one needed to be conceived for this application. The goal of the distance measure is to provide a measure of similarity between the activity patterns form different electrodes, while preserving the temporal detail, i.e., we were not interested in methods that involved time warping. Adherence to this strict time enforcement was important since the goal was to find the underlying canonical time courses that hint at the neural mechanisms of speech production and hence the detailed timing is important.

While we are interested in the similarity between activity patterns, not all aspects of the activity should be treated equal. Small differences in activity around baseline levels, i.e. around z-scored values of zero, should have negligible impact on similarity score. These differences in variation are not expected to be due to task processing as they are below measures of significance, either as measured through the Kalman filter (APPENDIX A) or with traditional statistical significant measures (i.e. bootstrapped t-test). Differences in variation when the significance threshold is crossed, on the other

hand, should be accounted for more within the distance measure. Previous studies have supported this rationale. In (Dichter et al., 2016) higher ECoG signal variability was present during non-task periods, with a reduction in signal variability during task specific time points, such as during stimulus onset. This shows that high signal variability exists during non-task periods, which has the potential to integrate to provide large distance values if not properly handled. Thus, we want to emphasize differences due to the task, while mitigating differences that are natural variations during the absence of activity.

We therefore developed a distance measure that puts less weight in differences during insignificant portions of the signal, around z-scored values of zero, and puts more weight in the significant portions of the signal. This helps prevent integrating many small differences in variation when neither signal is deviating from baseline activity and has grounding in prior observations of ECoG data and neural recordings and functionality. This also helps mitigate the need to heavily preprocess the signals to reduce the insignificant variation, like others have done.

We empirically form these properties into a measure by taking point-wise distances between a pairwise set of electrodes and then taking the square root of the sum of the squares of all the point-wise distances Equation 1. We use an exponential distance measure between the point-wise samples to give more importance weighting to significant parts of the signal. This measure is similar to the classical Euclidean distance, but the measure is in exponential space instead of linear. In CHAPTER II it was discussed that there is starting to emerge a trend of a shift in neural analysis to the use of nonlinear measures. We follow this trend and motivate its use over linear measures. For

our application, linear measures were found to integrate insignificant portions of the signal and in many of our analyses hid important features such as the differences between ramp and symmetric shapes. This measure is more formally presented and defined in CHAPTER III (§III.3.viii).

### *B.1.ii Clustering*

In this subsection, we elaborate on the hybrid clustering approach that was used. Our hybrid clustering approach combines elements from hierarchical and partitioning cluster methods. In particular, agglomerative hierarchical clustering and a modified k-centroid clustering hybrid algorithm is used. The distance measure just discussed in §B.1.i is used to assess similarity between electrodes activity and provide the measure used in the link function for grouping electrodes. The link function employs a centroid approach, in which the centroids (average activity from all electrodes in the current clusters) are compared when determining which clusters to merge. The partitioning method is used as refinement at each step, as described in the following.

We perform our hybrid clustering with cluster refinement through an iterative approach. Each electrode time course is initialized in its own cluster. Therefore, starting with the number of clusters equal to the number electrodes, as is the initialization of agglomerative hierarchical clustering. The distance between every cluster centroid, which at this point is just individual electrode time course, is computed. The two clusters with the smallest distance between them are merged. The new cluster centroid is updated to the mean of the member electrode time courses. At the next iteration the process is repeated. This continues until all electrodes are merged into a single cluster.

The process just described is agglomerative hierarchical clustering with a centroid method link function and the distance measure from §B.1.i. However, using a centroid method link function creates a non-monotonic cluster tree. Monotonicity is broken when a new cluster is created with a centroid closer (smaller distance measure) to a cluster that was not involved in the merge than the distance between the two clusters that formed the new cluster. To mitigate this and produce a strictly monotonic cluster tree, we employed a partitioning refinement step at each iteration. This step looks to partition the data based on the number of clusters for the step, i.e. *k* is determined by the current iteration which provides the number of branches in the cluster tree. The partitioning step works to refine the cluster assignments to ensure a monotonic cluster tree. Several algorithms were explored for the partitioning step, with k-centroids producing the best results. The pseudo-code for this approach is provided in Table 4.

**Table 4: Hybrid Clustering Pseudo-Code**

```
# Initialization
N = number of data points
K = N
C = compute_centroids(data points)
D = compute_pairwise_distances(C)

For iteration = 1:N-1

   # Number of clusters for iteration
   K = K - 1

   # Update – Hierarchical Step
   index = argmin(D)
   C = merge_clusters(C, index)
   D = compute_pairwise_distances(C)

   # Refine – Partition Step
   C_k-centroids = compute_k_centroids(C, K-1)
   method_index = argmin(min(D), min(D_k-centroids))
   if method_index == 2
      C = C_k-centroids
```

This hybrid approach ended up being similar just using agglomerative hierarchical clustering with a centroid link function. The main difference between that hybrid and agglomerative hierarchical methods is that the hybrid approach produced a monotonic cluster tree, while the sole agglomerative hierarchical clustering approach had non-monotonic points along the tree. However, the resulting clusters presented in the earlier chapters resulted from both methods. The agglomerative hierarchical method is a more established clustering method than our variant, but we maintain our hybrid method as our baseline approach since it produces a monotonic cluster tree that provides an easier way to determine the number of clusters.

Figure 27 shows an illustration of this clustering method to help with understanding the method. The figure illustrates a dendrogram view of the cluster tree (center). The dendrogram pictorially shows the cluster tree from left to right. On the left hand side each electrode is its own cluster and on the right everything is joined into a single cluster, where clusters are denoted by horizontal lines. As you move left to right, individual clusters get merged together. Cluster merges occur at the locations of the vertical lines, where two clusters (horizontal lines) from the left get combined to form one cluster (horizontal line) on the right of the merge. The distance moving left to right (along the x-axis) captures the distance measure between clusters that are merged. Thus, merges closer to the left of the plot occur between clusters that are most similar. This is illustrated by Channel A and Channel B in the figure being merged together relatively close to the left side of the dendrogram. Visually it can be seen that these channels have a high degree of similarity. Channel C is a bit more dissimilar from A and B and hence

gets merged with A and B further to the right, representing a greater distance measure needed for that merge. Selecting a distance measure for determining the cutoff between clusters that are similar and dissimilar is the topic of the next subsection, §B.1.iii.

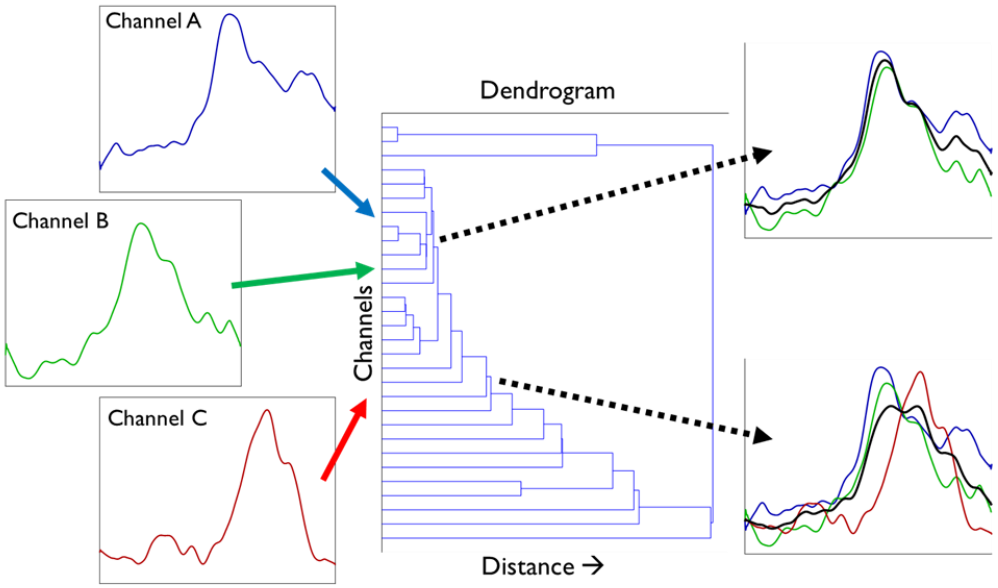


**Figure 27: Dendrogram View of Cluster Tree**

Dendrogram illustration of cluster tree (center) shows at what distance (horizontal displacement) each channel (vertically stacked) gets merged together. Moving left to right, each channel starts off as its own cluster and iteratively gets merged until there is one cluster containing everything. Channel A, B, and C temporal profiles are plotted for illustrative purposes to depict when these channels would be clustered together.

### *B.1.iii Selecting Number of Clusters*

The clustering methodology described in §B.1.ii creates a cluster tree where each step along the tree (iteration) provides the distance measure that was used to merge clusters at that step. Taking these distances over the number of clusters forms the cluster tree. As discussed in CHAPTER III, the goal is to find the correct step along the cluster tree to select the number of clusters for producing the characteristic activity patterns. This turns into selecting a tolerable distance for a measure of similarity, which we shall

refer to as the distance threshold. Several methods were analyzed for determining the distance threshold, many of which yielded the same results.

A good starting point when performing this type of determination is the "elbow-method". The elbow method looks at where there is an elbow in the cluster tree, hence the importance of having a monotonic result. An elbow can be thought of as the point along the curve where there is a noticeable increase of the distance threshold needed to move to fewer clusters. There is no golden rule for what constitutes an "elbow", but visual inspection provides a coarse rule of thumb that helps to guide the selection of analytic measures and provide a ballpark for where a tolerable distance threshold should be set. In the results contained within the chapters of this dissertation the visual elbow lined up with the analytic methods described next. This was not the case with some of the alternative clustering methods described in §B.2 or alternative preprocessing discussed in APPENDIX C. The elbow method was used to provide a sanity check and corroborating evidence.

To set our distance threshold, we looked at the percent of variance explained by the clusters. This is discussed in CHAPTER III (§III.3.viii) and laid out in (Goutte et al., 1999). This approach looks at the measure of variance explained by the clusters within each step of the cluster tree, with the difference in moving from one step to the next providing the additional variance that is explained by adding an additional cluster. This approach also requires setting a threshold, but the units are now easier to understand – percent variance explained. Using this measure, we were able to see a quick flattening of the additional variance explained by going to additional clusters, with minimal

explanation resulting from moving beyond the distance threshold that was selected. This approach also needs monotonicity for correct interpretation, hence our work to ensure a monotonic cluster tree. So, while we cannot fully get away from using a heuristic, we are at least able to move towards a distance threshold decision that provides intuitive and interpretable understanding.

It was touched upon in §B.1.ii that the results of a sole agglomerative hierarchical clustering method with a centroid method link function produced the same results as the hybrid approach, despite it producing a non-monotonic cluster tree. In that approach, points in the cluster tree that break monotonicity result in cluster tree locations where moving from more to fewer clusters decreases the distance threshold that needs to be set, instead of increasing it. Likewise, it also means that adding an additional cluster would decrease the percent of variance explained or, conversely, using fewer clusters would increase the percent of variance explained. This forms a discontinuous location within the cluster tree when analyzing how many clusters to select, as points in the tree would never be selected no matter the threshold. Hence, these locations can be removed, or pruned, to create a monotonic cluster tree.

This cluster tree pruning was used with the sole agglomerative clustering to generate a monotonic cluster tree. This gives the same result as the hybrid approach, with the same number of clusters and same electrode assignment to the clusters. This was, however, only true at the region of the cluster tree that was the focus of the selection criteria. It was not the case if the desire was to increase the percent of variance explained by fractions of a percent and accept more clusters. The clustering methods diverge in

their results slightly in this case. We were not interested in minute increases in percent variance explained and instead wanted to find a parsimonious representation that captured the characteristic shapes of speech. Thus, both methods were interchangeable at our distance threshold.

### *B.1.iv Conclusions*

The hybrid method ended up not being necessary as the special handling of the agglomerative hierarchical clustering produced the same results. We discuss the hybrid method here and use it as our main form of clustering due to a lot of the alternative distance measures tested, as the hybrid method produced far more stable results in a lot of those cases, but the results presented within this dissertation also were produced from the simpler, and more established, agglomerative hierarchical clustering method.

## B.2 Alternatives

This section contains an overview of some alternatives explored, including the distance measures and clustering approach. In particular, this section serves to note the rationale behind looking into these alternatives and details on why they were not sufficient. This section will provide brief descriptions touching on several aspects of the clustering process, including the distance measure (§B.2.i), preprocessing (§B.2.ii), time-series representations (§B.2.iii), clustering method (§B.2.iv), selecting the number of clusters (§B.2.v), and cluster assignment (§B.2.vi). Particular focus will be put on the distance measure (§B.2.i), as that was one of the novel contributions of this work.

### *B.2.i Distance Measures*

As discussed in the introduction to this appendix, there are many different options for distance measure, see (Warren Liao, 2005; Serrà and Arcos, 2014; Aghabozorgi et al., 2015) for a review. We ultimately settled on an exponentially weighted distance measure to preserve time and emphasize elements of the ECoG temporal profiles that were further away from z-scored values of zero, i.e. significant parts of the signal. If we relax the emphasis on these properties, there are several other measures that could be considered.

Similar to the majority of the research in this area we initially started with a linear distance measure before focusing in on the nonlinear exponentially weighted distance measure. We initially used the Euclidean distance measure (Anton, 2010), as laid out in Equation 10 where *x* and *y* are time courses from separate electrodes and *N* is the total number of samples in the trail. This distance is one of the most common, and evaluates distance in linear space. It was found to be too sensitive to slight variations in differences around z-scored values of zero, i.e. no significant activity. This sensitivity integrated over the duration of the trial to have a significant impact on the measure of distance.

**Equation 10: Euclidean Distance Measure**

$$d(x,y) = \sqrt{\sum_{n=1}^{N}(x_n - y_n)^2}$$

This brought about our motivation to de-emphasize activity close to z-score values of zero and its integration potential. This integration potential is illustrated in Figure 28 where two time courses (blue and green) can visually be seen to be similar. However, the right side of the figure shows the difference between these two signals, as

illustrated by the shaded yellow area. This results in a large amount of distance accumulating in the beginning of the signals when using a linear measure.

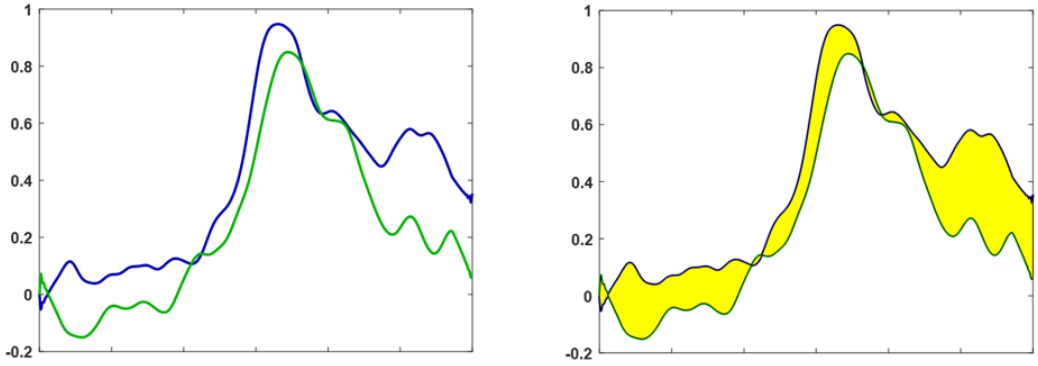


**Figure 28: Linear-Space Integration of Non-Significant Activity**

Left: Activity patterns from two different electrodes, blue and green. Right: Distance between the two electrodes shaded in yellow. In linear-space the point-wise differences in the beginning of the signal can be seen to have a major contribution to the overall distance measure for the two electrodes, constituting over 1/3 of the total distance.

As a contrast, Figure 29 shows a notional depiction of the difference in emphasis for point-wise differences between signals from the Euclidean (left) and exponential (right) distance measures. The x and y scales capture the individual sample z-score values for the two time courses being compared, respectively. Color represents the emphasis, or weighting, that the point-wise differences provide to the overall distance measure, with blue being low and yellow high. In the Euclidean space (left plot), it is seen that equal weighting is given to the same differences no matter what the z-score values were. Thus, the emphasis is indifferent to where you are in the activity pattern. Differences when the z-score values from both signals are close to zero (bottom left of figure) are given the same emphasis as similar differences when both z-score values are high (top right of figure). The exponential distance measure (right plot) provides varying levels of emphasis dependent on where you are in the activity pattern. When both values

are close to zero (bottom left) there is a broader region that expands out de-emphasizing these differences. Contrast this with when both values are high (top right). Here smaller differences in the activity patterns start to add more emphasis and therefore contribute more to the distance measure.

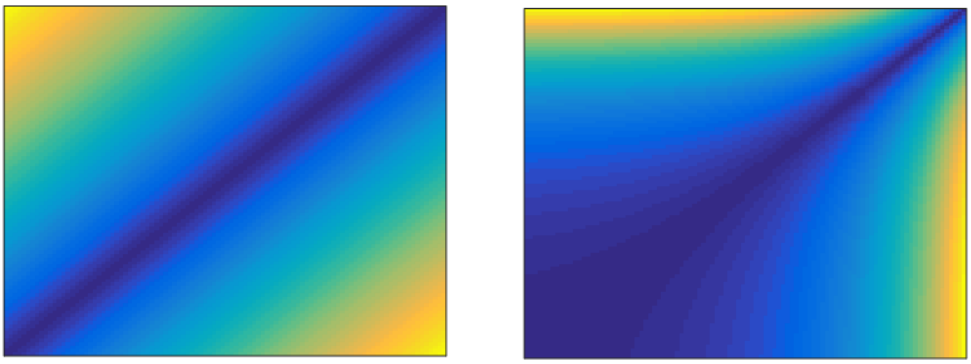

**Figure 29: Impact of Linear vs Nonlinear Distance Measures**

Plots illustrate the notional emphasis put on the distance between two electrode activity patterns as a function of their z-score values. X-axis represents z-scores for one electrode, with a value of zero at the origin and larger values moving to the right. Similarly, the y-axis represents z-scores for the other electrode, with a value of zero at the origin and larger values moving up the axis. To determine the emphasis of point-wise distances, the location of the intersection of the z-scored values for each activity pattern is used. Blue denotes lower relative emphasis and yellow represents higher emphasis. The diagonal represents points where the z-score for each activity pattern is equal (dark blue), and hence there is no difference. Left: Euclidean distance measure. Right: Exponential distance measure.

Another alternative is cross-correlation (Grami, 2019). The cross-correlation function measures the similarity between two signals by holding one constant and measuring the similarity of the other at different time lags in reference to the first. At each lag, the similarity is measured as in Equation 11. $R_{xy}$ is the cross correlation between the two signals, $x$ and $y$. It is a function of time lag, $t$, and therefore can provide a distance measure between the two signals not just while they are aligned by the task time, but also with temporal shifts.

**Equation 11: Cross-Correlation Distance Measure**

$$R_{xy}(t) = \begin{cases} \sum_{n=0}^{N-t-1} x_{n+t} y_n, & t \geq 0, \\ R_{xy}(-t), & t \leq 0. \end{cases}$$

To ensure that the distance measures across the different pair-wise comparisons of electrodes were the same, the cross-correlation results were normalized so that the distance of the autocorrelation (cross-correlation between an electrode temporal profile and itself) at a lag of zero was one. This is shown in Equation 12, where $R_{xy}$ is the cross-correlation from Equation 11.

**Equation 12: Auto-Correlation Function**

$$\frac{R_{xy}(t)}{\sqrt{R_{xx}(0)R_{yy}(0)}}$$

The cross-correlation at a lag of zero, $R_{xy}(0)$, provides an estimate of the similarity of two electrode time courses when time is preserved for both signals to be time-locked to the task. This was the main measure that was evaluated. Time lags not equal to zero, $t \neq 0$, were also evaluated to see if an electrode's activity was similar to a shifted version of another electrode's activity, i.e. the shape was the same but there was a shift in time. Early analysis used this approach to look for propagation of canonical characteristic activity through the speech network, i.e. time delay versions of the same shape at the various steps of processing. However, the underlying mechanics of this method are rooted in linear space and thus bound and limit the insight that can be found.

If we go a step further in relaxing our constraint on time, dynamic time warping (DTW) allows for a different form of comparison (Berndt and Clifford, 1994). DTW

measures temporal similarity between time series that may differ in speed, hence the allowance of time warping. This temporal nonlinear warping allows for differing acceleration and deceleration between similar activity patterns, but mandates that the general shape must still be present. Each sample from an electrode is lined up with corresponding sample(s) from the other electrode, potentially at different time indices. The overall alignment must maintain the ordered time nature - aligned indices need to be monotonically increasing. This results in having points in the alignment where some samples from one electrode must be repeated while waiting for the other one to catch up, i.e. a deceleration. At other times, multiple samples from one electrode get mapped to a single sample of the other to quickly get back to alignment, i.e. an acceleration. Refer to (Keogh and Ratanamahatana, 2005) or others for a more detailed review.

Similar to cross-correlation, DTW also allows for exploring how characteristic activity patterns may propagate through the brain when relaxing the time requirement. Some earlier research of ours went down this path. We used the Keogh lower bound (Keogh and Ratanamahatana, 2005) as our distance measure, rather than the actual DTW distance measure. This was done for simplicity as the Keogh lower bound is a fast implementation for computing the DTW. This measure was compared to DTW on smaller sample sizes and final results to ensure using the approximation did not generate unwanted behavior.

Taking another approach to distance, we also explored performing transformations of the time series data before assessing distance. One of the common transformations for time series data is going to the frequency domain, which we explored

to see if there were findings within the frequency domain that were not readily visible in the time domain.

In particular, we transformed the time series into the frequency domain using Fast Fourier transforms, FFT (Cooley and Tukey, 1965). To enable differences in timing to be more important than differences in amplitude, or vice versa, we set a factor, $\rho$, that traded weighted the magnitude component of the Fourier representation. First, each time-series, $x$, is converted to the frequency domain using the FFT and normalized by the number of elements in the sequence, resulting in a frequency domain representation $F_x$. This normalization term could be dropped since all sequences had the same number of elements, $N_x$, but is included for completeness in Equation 13.

**Equation 13: Frequency Domain Representation of Time Course**

$$F_x = \frac{FFT(x)}{\sqrt{N_x}}$$

Next, a function, $G_x$, is taken to project the weighted magnitude of each Fourier representation of the signal, $F_x$ and $F_y$, under the phase of one of the signal Fourier representations, i.e., $F_x$. The projection is also taken with the signals swapped, $G_y$, where the x and y components in Equation 14 are switched. The * denotes element-wise multiplication in Equation 14.

**Equation 14: Weighted Magnitude Projection by Phase**

$$G_x = \left((0.5 + 0.5\,\rho)|F_x| + (0.5 - 0.5\,\rho)\left|F_y\right|\right) * e^{i\,angle(F_x)}$$

The measure of similarity, or distance measure, $d$, is then the L2 norm, or Euclidean norm, of the difference between these two projections, as shown in Equation 15.

**Equation 15: Frequency Transformation Distance Measure**

$$d = \left\| G_x - G_y \right\|_2$$

The factor, $\rho$, becomes important in how much weight to put into the magnitude difference between the two signals. A $\rho$ value of 1 reduces this expression to become the norm of the difference in the *x* and *y* time series expressions. A $\rho$ value of 0 makes the distance measure independent of the difference in the magnitudes of the frequency response, i.e. $|F_x| - \left|F_y\right|$.

Several other methods were tested, but fell within similar buckets for rationale as those just explained: a) methods that preserved time and computed sample-wise distance measures (such as the Euclidean method or the used exponential distance measure), b) methods that allow for time shifts (such as cross-correlation), c) methods that allow for time warping (such as the acceleration changes allowed in DTW), or d) methods that explore transformations that either move the measure away from a temporal representation (such as the transformation to the frequency domain) or parameterize it through a model (not discussed here as most these methods remove too much important information for understanding the time course of human speech). All of these methods try to estimate a similarity (or difference) between electrode activity patterns, but do so in slightly different ways and thus have slightly different interpretations.

We are not unique in applying clustering analysis to look at ECoG during speech, refer to CHAPTER II for a review (§II.3). The methods used by these other groups, however, rely on linear distance measures for assessing similarity between time courses.

One group in particular, the Chang Lab at University of California, San Francisco, has been exploring clustering analysis methods, e.g. see (Leonard et al., 2019). One of the techniques that this lab often uses is factor analysis. This is a statistical technique that is based on the correlation matrix to identify interrelationships between many variables. The interrelationships are captured in the form of factors, which capture the common variance in the data and discards the unique variance, hence relying on the inherent correlation in the data to provide a reduced set of representations of the data, factors, that compactly describe the data. These factors are linear combinations of the data, which in our type of analysis is linear combinations of the electrode temporal profiles. In particular, the Chang Lab has used non-negative matrix factorization (NMF) (Lee and Seung, 1999) and a specialized version of this called convex NMF (Ding et al., 2010), which has nice properties for clustering (Ding et al., 2005). This results in a soft clustering algorithm whose factors are the cluster centroids and electrode to cluster assignment comes from linear combinations of the factors, hence soft clustering which is discussed in §B.2.vi. Further, many of these alternative methods are validated with a partitioning method, such as k-means which is often used by the Chang Lab. This again is based on correlation and again reliant on a linear distance measure.

In some of our analyses we enforce additional constraints on the distance measure. In particular, in CHAPTER IV we expand the distance measure to encourage spatial similarity between electrodes. The distance measure is updated to be the combination of a temporal and spatial component. The distances just discussed were also explored in this context for the temporal component.

Several alternative options were also explored for the spatial component, §IV.2.ii. Different parameterizations of the radial basis function (RBF) kernel were explored, as well as other RBF forms. This included removing the bilateral flattening, as discussed in CHAPTER IV (§IV.2.i), and therefore separating clusters that span hemispheres. Kernels not set in a radial basis were not considered due to complexity. Similarly, kernels that were dependent on where they were cortically centered were not considered for complexity reasons. These types of kernels would be dependent on cortical locations and would take into consideration the neural pathways of the brain, e.g. a kernel centered in a specific Brodmann area may put more weight to other locations within the same area versus moving to neighboring Brodmann areas.

How the two components, temporal and spatial, of the distance measure from CHAPTER IV were combined was also an active area of exploration. The final solution itself had two parts, Equation 8 (§IV.2.iii). The first component was only based on the temporal component, while the second was a mixing between both the temporal and spatial components. It was deemed that the temporal component was the most important and thus made up the first part by itself. The second part was added to this to bring in the influence of the spatial component. This second part was constructed as a mixture of the temporal and spatial components. The mixture was a multiplication of the temporal and spatial distances, scaled by a factor. This was experimentally found to perform the best in terms of the correct balance between temporal and spatial importance.

Other things explored were direct mixing (i.e. dropping the first part of Equation 8 and only using the second part) and a weighted linear combination (scaled addition of

the temporal and spatial components). The scaling factors that balance the importance of the temporal versus spatial component was found to be the most sensitive parameters of all of these functions for generating resulting clusters. Across the set of functions tried, it was found that more weight should be put in the temporal component so that the distance measure did not reduce down to generating anatomical brain parcellations and was able to generate meaningful characteristic activity patterns. Too much weight in the spatial component was found to produce a set of symmetrical activity patterns that had larger variances and were mostly separated by their times of activation and spatial locations. The unique ramp shapes got washed out in these cases.

### *B.2.ii Preprocessing*

Several variations on preprocessing were tested, one of which is discussed here. In CHAPTER III, normalization was applied to each electrode time course so that the signal was scaled to be between zero and one. This results in all electrodes having activity that is normalized to have the same power range, enabling characteristic activity patterns that are scale invariant. This puts the focus on the temporal dynamics, i.e. the shape of the activity pattern, instead of the amount of power contained within the activity. Earlier experiments looked at what would happen without normalization by looking at how clusters would form from both the shape and amount of power contained within the activity. Variance was found in the power ranges of the electrodes making it difficult to form meaningful clusters over the small subject sample size.

### *B.2.iii Time Series Representation*

The temporal profiles used to perform the clustering in the chapters resulted from the minimally preprocessed high gamma representation of the ECoG data. This was desired, as it reduces the number of preprocessing steps involved and hence the steps that potentially remove, transform, or alter the neural activity. However, additional preprocessing was explored to see if there was a tradeoff between additional processing that potentially removes information and a better set of resulting clusters. One of the goals of this additional processing was to reduce dimensionality for the clustering step, and hence reduce the complexity of the analysis. Another goal was to smooth the data to capture the general shape, but remove minor deviations that existed in the signals. Smoothing helped to make shapes with fewer dynamics, but also slightly shifted the time-course and activity pattern of the signals, so was not desired. However, smoothing did help create better clusters prior to moving to an exponential distance measure as it helped to mitigate the integration effects of small deviations during insignificant portions of the activity as discussed in §B.1.i. Once the exponential distance measure was tested it out-performed smoothing methods.

Methods used to perform this complexity and dimensionality reduction by smoothing the data fall within a class of techniques called model-based. In model-based techniques there is an explicit model for the signal and the goal of the technique is to fit the model's parameters to capture and describe the signal. Several of these techniques will be discussed, including principal component analysis (PCA), wavelet transform, and autoregressive (AR) models. Additionally, the method from APPENDIX A will also be

explored as an alternative representation. Other techniques from this class were also considered. Some of them were implemented, such as splines, but they provided poor fit and generalization. Their primary pitfall was their need for fine tuning that did not hold from one condition to the next. Thus, they lacked the ability to generalize and became point solutions that lacked evidence of holding up in other tasks or subjects. These techniques will not be discussed further.

Principal component analysis (PCA) is a commonly used technique to capture the underlying trends in a data set, while reducing the dimensionality (Pearson, 1901; Hotelling, 1933; Jolliffe, 2002). It is a data-driven technique, in that it does not rely on any basis method, but instead on the data itself to generate the principal components. Many variants of PCA exist, but in our implementation we stuck to the multivariate normal (Gaussian) distribution that is used for inferential purposes.

PCA seeks to find a linear combination of the variables to create new signal representations that have fewer variables with maximum variance, with the goal of maximizing the variance to form a representation that best explains the original signal with a reduced set of variables. These fewer variables, or principal components, are orthonormal. Thus, PCA seeks to take a dataset that has potentially correlated variables in it and reduce it to a smaller dimensional representation with uncorrelated principal components. PCA itself is a form of factor analysis, which was discussed in the context of distance measures in §B.2.i, where the factors in PCA are based on the total variance.

In our implementation, the different electrode time courses become the variables and PCA is used to find the combination that “clusters” these temporal profiles into a

smaller set of principal components that contain most of the variance across all electrodes. This is a form of soft clustering, where each temporal profile can be a combination of multiple principal components. However, we did not seek to use PCA for clustering, but rather for preprocessing. Thus, we performed PCA to get the principal components and then used them to construct signal representations for each electrode from its linear combination of the principal components. This results in a smoother version of the original time course.

Results varied using this method. Many of the smoothed time courses captured the basic shape and time structure of the original signal, which was what was hoped for. However, a nontrivial subset of the smoothed temporal profiles got manipulated in a way that the overall shape lost much of its original structure and was not a good smoothed version of the original signal.

Many of the model-based methods, including PCA, assume that the data is stationary. Speech and the neural mechanisms that produce them are dynamic in time, frequency, and space. We also looked into models that were more representative of this non-stationarity. Extensions have been made to stationary model based techniques, such as functional PCA (Jolliffe, 2002; Jolliffe and Cadima, 2016) to allow for time evolution. We instead sought to explore a technique that had more dynamics built natively into it, seeking a method that can describe the signals in time and frequency simultaneously and properly handle variability. Wavelet-based methods have these features and were another approach we pursued for transforming the representation of the data prior to clustering. More information about our wavelet approach can be found in §C.2.iv, where we discuss

using it as a filtering step in our preprocessing stack. Results were similar to the other techniques explored that did a form of smoothing.

Another technique that others have used is an autoregressive (AR) model. This is also a technique that we explored for filtering and a brief discussion of it can be found in that section, §C.2.v.

Lastly, we also explored using the learned Kalman filter estimates from §III.3.ix and APPENDIX A as the representation used for clustering. The additional preprocessing through the Kalman filter performs smoothing like the other alternatives described, but does so in a different way. Here, the smoothing emphasizes the trends present in the activity. The trends are broken up by change points and are relatively linear in each segment between the change points, due to the form of the Kalman filter. The change points create some discontinuities, which is undesirable for clustering as they create discrete temporal sensitivities. However, reducing the signal into segments that capture cleaner versions of the underlying linear trend is attractive for clustering. The segments maintain the important time points of the signal and the general structure of the original shape, making it easier to assess similarity between different electrodes. The discrete change points make it easy to line up points of significant trend inflection when the same significant times exist across activity patterns, i.e. time of peak.

Many nice properties came out of this, but the slight timing differences at change points sometimes summed up to large differences in a distance measure and were overpowering. This sensitivity was ultimately the limiting factor when trying to find a set of characteristic time courses. The Kalman filter approach worked very well, and in

fact outperformed many of the other techniques, when large number of clusters were considered with limited electrodes in each one. However, performance quickly dropped off when trying to get to a smaller set of representative clusters as slight time differences (activations, peaks, etc.) caused poor cluster merges. To mitigate this, cluster centroids could have been constructed from the ECoG high gamma signals and the Kalman filter only used in the link function. This mitigation was not explored as it began to add more steps, creating a complex model that was becoming harder to intuitively understand.

In addition to the model-based steps described, a second class of preprocessing was also tested. This second class of preprocessing does not try to model the signal to get a new representation for clustering, but instead use subsets of the signal for clustering. All the techniques discussed thus far have used the entire trial period, with different techniques to emphasize different parts of the signal. Here, instead of figuring out how to emphasize different parts of the signal we explicitly ignore parts of the signal and only use subsets of the signal that were significantly responding to the task, i.e. significantly deviating from the baseline period or a z-scored value of zero.

This was only explored briefly, as challenges arose in how to assess similarity between two signals when different sub-signals resulted in different parts of the trial that were used that only partially overlapped, with no overlap in the extreme. One attempt set periods of the signal that were not significant to zero. This has the effect of maintaining a signal representation throughout the whole trial period, with perfect similarity when neither electrode has significant activity. However, this created differences that significantly integrated distance at points where one signal was significant and the other

was not, and therefore set to zero. Distance measure approaches, including the exponential distance measure, did not fully help to mitigate this effect. This approach ended up functioning to put a lot of weight on the slightly significant parts of signals and de-emphasized the largely significant periods of the signal, including times like peak activity. This was the opposite of the intention of the approach.

Another attempt to mitigate this was to remove non-significant periods from the analysis altogether so they do not contribute positively or negatively. This created the challenge noted above with how to deal with computing a distance at a time sample when one signal is significant, and hence its value should be used, while the other is not, and hence its value should not. Several attempts were made to try to address this, with nothing producing tolerable results and hence sub-sequencing was no longer pursued.

### *B.2.iv Clustering Methods*

A hybrid clustering approach was ultimately used for the results of this dissertation. Alternatively, an agglomerative hierarchical clustering approach could have been used and resulted in the exact same results, as covered in §B.1.ii. Alternatives pursued included variants of agglomerative hierarchical clustering with various link functions and combinations of distance measures. Additionally, partitioning clustering methods were also explored, including k-means, similar to what others have done to validate their methods, i.e. (Leonard et al., 2019). Some additional custom iterative clustering techniques were also tested, but as research progressed it was determined that most of the difference in performance was a function of the distance measure used rather than the clustering method employed. Thus more time was devoted to refining the

distance measure, which in turn produced similar results across methods. Thus, choice of clustering method does not seem to play a major role in the results found in this work.

### *B.2.v Determining Number of Clusters*

CHAPTER III describes the details on how the number of clusters was selected from all possibilities in the cluster tree. Figure 30 illustrates what a cluster tree looks like to help build some intuition. Here it can be seen that the distance threshold, which was covered in §B.1.iii, is clearly a function of the number of clusters. Setting a low distance threshold results in many clusters (right side of figure). Conversely, setting a high distance threshold results in fewer clusters (left side of figure). The red marker highlights the point at which the percent of variance explained only gets slightly higher with the addition of more clusters. This is the point at which the number of clusters to use is selected. It is also visibility seen to be the location where the “knee in the curve” exists, showing why this general rule of thumb still holds a lot of power for selecting the number of clusters.

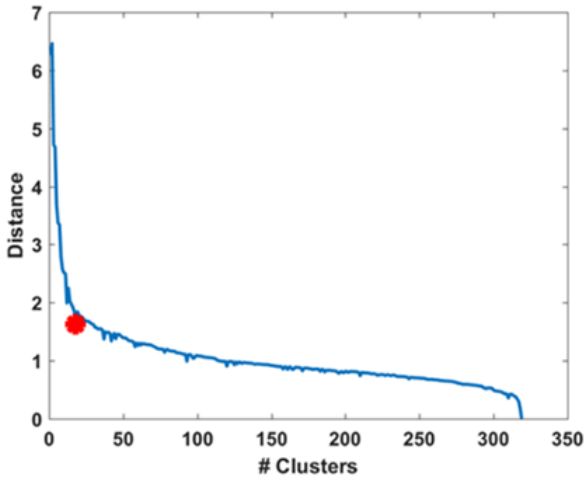


**Figure 30: Example Cluster Tree for Selecting Number of Clusters**

Cluster tree (blue) is shown as the distance needed (y-axis) to result in a specific number of clusters (x-axis). The red marker denotes the location that is selected for providing the number of clusters using the percentage of variance explained, which sets the distance threshold for selection. This visually aligns with the "knee in the curve" heuristic.

Many different techniques exist to choose the number of clusters, with no clear choice in any situation. Two example alternatives are the Akaike information criterion (AIC) and Bayesian information criterion (BIC). Both of these techniques penalize using more clusters than you need to in order to promote more parsimonious models. Many other options exists, refer to (Warren Liao, 2005; Aghabozorgi et al., 2015) for more complete coverage. Other options were not explored in detail as the "knee" ended up being fairly pronounced in our results, thus providing strong evidence backing the percent variance explained technique that we used.

### *B.2.vi Cluster Assignment*

Finally, a quick note on cluster assignment. All of the clustering methods explored in this work performed hard assignment, in that electrodes are only assigned to a single cluster. Soft clustering on the other hand would have allowed electrodes to be

assigned to multiple clusters. This partial assignment typically comes in the form of a weight that determines how much electrodes fit within each cluster.

It was determined early on that we would only pursue a hard clustering approach and not soft clustering. This was due to the desire to find canonical activity patterns that characterize speech. In a hard clustering approach, an electrode time course needs to match a cluster centroid over the entire duration of the trial, thus capturing the underlying neural dynamics throughout the speech processing that is taking place. In a soft clustering approach this need not be the case. You could have an electrode time course that is matched to two separate clusters, where it matches one cluster in the activity ramp up stage and another cluster in the activity decay stage. This breaks up the notion of a canonical activity pattern as the underlying dynamics of that electrode are now captured by multiple clusters, but neither completely captures the neural dynamics. This would prevent the differentiation of ramp and symmetrical shapes, as it was found that these clusters often share parts of the trend, as illustrated in Figure 11 (§III.6.iv). Additionally, soft clustering often produces non-unique partitions, meaning that multiple different activity pattern sets could be constructed that could equally describe the data.

Even with soft clustering you still have the burden of cluster assignment. If an electrode fits 100% in a cluster, you would assign it to that cluster and can describe the electrode location as involved in the activity pattern of the cluster. What if the electrode is only 50% assigned to the cluster? Do you then say it is only 50% performing that part of the cluster activity pattern? What about 10% or 1%? And what does this mean in terms of cortical involvement? In order to make the results interpretable in regards to the

task being studied some threshold needs to be set. Electrodes could contribute to multiple clusters, but having them contribute to too many will not provide insight into the underlying dynamics. The argument in the past is that soft clustering allows you to capture multiple task processing functions that an area is performing. The argument extends to say that since neural data is heterogeneous we should allow for multiple cluster assignments to capture this heterogeneous underlying neural mechanics, e.g. (Leonard et al., 2019). We agree that neural data is heterogeneous, but we argue that if brain areas are truly performing multiple tasks, they will form their own cluster that captures this and is separable from brain areas that only perform one of the tasks.

# APPENDIX C: More on Methods

In this appendix, we lay out some additional information in regards to the methods. In doing so, we provide additional context, rationale, motivation, and methodology beyond what was discussed in the earlier chapters, with a primary focus on the content of CHAPTER III, which lays out the core methodology used in this research. In motivating the directions taken and providing rationale, we will also discuss some of the alternatives considered and the earlier approaches tried in this research. This appendix is not intended to lay out the full cohesive processing flow, but instead to hit upon some of the more critical elements.

Much of the preprocessing steps we used are common in electrocorticography research, with many researchers converging around a few common methods. In CHAPTER III we tie our methods back to those used by others. These common methods will not be discussed in detail here. Further, this appendix will not cover material involving the Kalman filter or clustering. More detail on those two aspects of the methods can be found in APPENDIX A and APPENDIX B, respectively.

## C.1 Methodological Flow

First, we orient to the entire processing chain by providing an overview of the methodological flow. This high-level description is intended to provide intuition and summary level understanding of the approach. This description is not meant to provide enough details to understand each element, but instead to compactly tie everything together. Figure 31 illustrates the processing flow, with a description following.

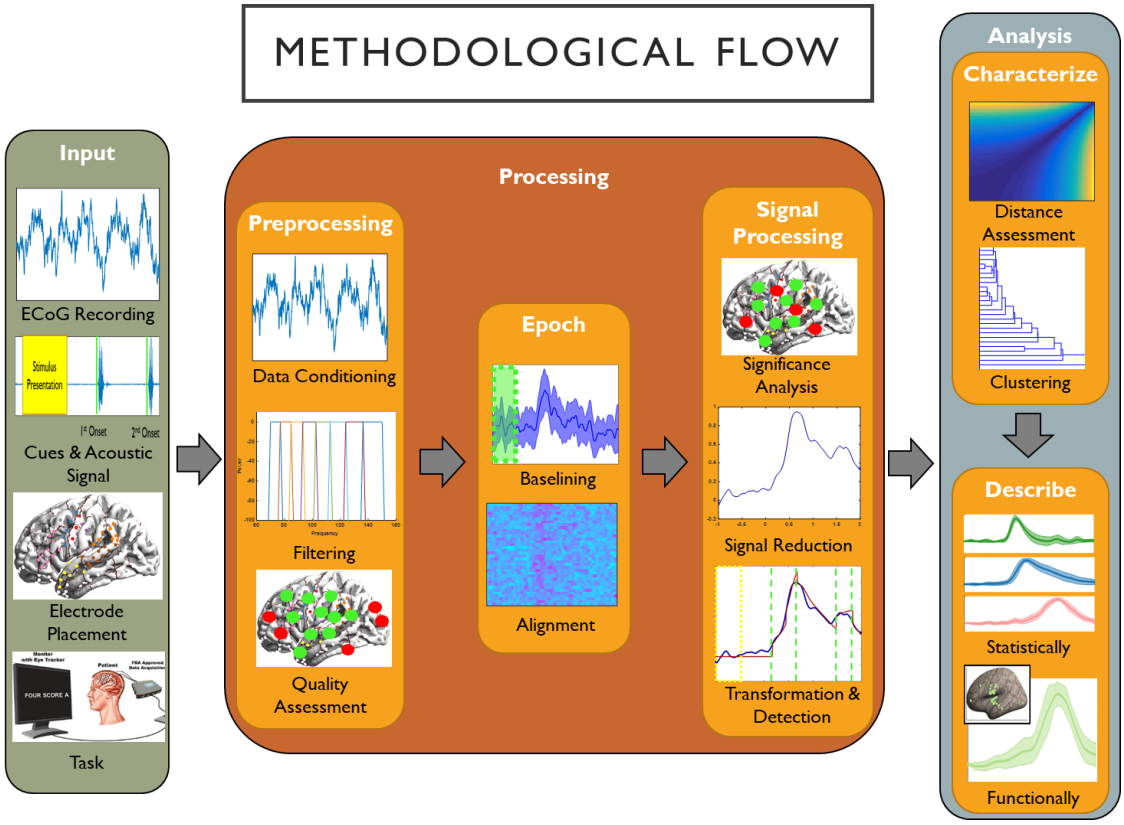


**Figure 31: Primary Processing Pipeline**

Methodological flow depiction of primary processing pipeline for work of this dissertation. Input data sources (left) are aligned, compiled, and manually corrected. Processing (middle) proceeds through several steps: preprocessing, epoching, and signal processing. Analysis (right) is conducted on the results to characterize and describe the data.

The input data sources, left side of Figure 31, consist of the electrocorticographic (ECoG) recorded data, acoustic response signal, electrode locations, task description, and cues and stimuli signals. Prior to running any processing, these multiple sources first need to be aligned. Semi-manual voicing onset and offset is conducted to mark the speech trials within the acoustic signal and remove any non-speech artifacts. All signals are then aligned to common reference points.

Data processing comes next, with three primary stages: preprocessing, epoch, and signal processing. In the preprocessing stage, the data is further refined through both

manual and automatic quality assessment to remove electrodes that have artifacts. The ECoG signal is then downsampled, cleaned (direct current and line noise removed), and re-referenced. Next, the data was filtered to the frequency band of interest. After this, the next stage is to epoch the data. Data for each electrode is aligned to a trail condition (stimulus presentation or voicing onset) across all trials and then averaged together to form an event-related spectral perturbation (ERSP). The signal processing stage follows. ERSPs are analyzed for significant task-related activity. Those showing activity are kept and further downsampled to reduce dimensionality for the analysis. Lastly, additional transformations and trend detection are performed for some variants of the analysis.

The analysis steps are used to characterize and describe the final results. The characterization step performs clustering to form characteristic time courses. Descriptions of the characteristic shapes are provided in the form of describing the temporal structure (trend analysis) and connecting it to anatomical and functional meanings. Next we turn to some of the important elements of these steps.

## C.2 Filtering

Using a Hilbert transform to extract high gamma frequency power from ECoG was the primary form of filtering used in this work. Here we provide a more complete discussion of filter options, including spanning into other frequency bands as well as different forms of filtering. Our ultimate decision to use the Hilbert transform was to best align with the vast proportion of prior studies that have used it. We also did not see any benefit produce by using a different method, so conformed to what others have done.

ECoG signals have been shown to follow a power law scaling with broadband amplitude changes indicating neural activity (Miller et al., 2009), have shown high gamma power phase-locked to theta oscillations (Canolty et al., 2006) and to broadband power in the motor cortex (Miller et al., 2012), and have shown high gamma tracking of speech acoustic envelops in the auditory cortex (Kubanek et al., 2013). Thus, filtering across several different power bands may provide additional insight into the neural mechanism and help to provide insight in the understanding their contribution to speech.

First, a brief discussion of the four different frequency bands that were considered is be provided. The four frequency bands are described as follows:

- ***Raw Signals*** *(no filtering)*: The raw signals were looked at directly to reduce the amount of preprocessing. This allows for analyzing activity contained across the entire frequency range. Early preprocessing steps were still used to clean up the data, i.e. removing line noise, but no further filtering was performed.
- ***Broadband*** *(4 – 250 Hz)*: Broadband activity has been found to link local field potentials (LFP) and ECoG activity with functional magnetic resonance imaging (fMRI) bold responses (Ojemann et al., 2013).
- ***High Gamma*** *(70 – 150 Hz)*: High gamma frequency band activity has been found to correlate with local neural activity (e.g. (Miller et al., 2009; Ray and Maunsell, 2011)) and will be the primary focus of study.
- ***Beta*** *(15 – 30 Hz)**:*** Beta activity has been found to be suppressed with local activity during a task (Pfurtscheller and Lopes da Silva, 1999).

Table 5 summarizes the frequency bands considered and their frequency extent.

**Table 5: Electrocorticography Frequency Bands**

| | Frequency (Hz) |
|---|---|
| Raw Signal | |
| Broadband | 4 – 250 |
| High Gamma | 70 – 150 |
| Beta | 15 – 30 |

There are several methods for filtering the raw data to get the band responses noted above. We primarily focused on the Hilbert transform, but explored other methods to see how results differed. The following subsections detail each of these alternatives, which consist of 1) Hilbert transform, §C.2.i, 2) highpass and lowpass, §C.2.ii, 3) multi-taper, §C.2.iii, 4) wavelet, §C.2.iv, and 5) autoregressive model, §C.2.v.

### *C.2.i Hilbert Transform*

The Hilbert transform (Hupert, 1965) has been used for spectral analysis across different domains. Within ECoG research, the log-analytic amplitude of the Hilbert transform has been used to bandpass the ECoG recordings into band specific regions (Edwards et al., 2009; Chang et al., 2011; Ray and Maunsell, 2011; Moses et al., 2016). For high gamma, this is done using 8 logarithmically spaced bands across 70 – 150 Hz, such as in (Hamilton et al., 2018). The logarithmic scale is used due to its spectral properties (Gasser et al., 1982). The absolute value of the signal from each filter band is taken to get the power envelope. This is then averaged to get the high gamma power. Many high gamma analyses have used this approach since it was first proposed by

(Edwards et al., 2009). In the broadband power range, 4 – 250 Hz, this has been done using a filter bank of 42 center frequencies.

More formally, the bandpass filter for the Hilbert transform is designed as follows. For each logarithmically spaced frequency band a separate bandpass filter is designed using a Butterworth filter (Butterworth, 1930). The filters are symmetric around the passband and have 1 Hertz (Hz) of separation between their pass and stop frequencies, with 60 decibels (dB) of attenuation. The passband has been designed to have a ripple of less than 1 dB. The sampling rate of the signal that the filter was applied to was 1000 Hz. The signal is zero padded on each side and is filtered using a zero-phase forward and reverse digital filter to execute the Butterworth design using the *fdesign* and *filtfilt* functions in MATLAB (MathWorks, Natick, MA). The zero padding is then removed and the procedure continues as in the preceding paragraph.

### *C.2.ii Traditional Highpass and Lowpass Filters*

Traditionally, bandpass filters have been constructed using a highpass and lowpass filter to get the spectral content, such as with the fast Fourier transform (FFT) (Potes et al., 2012). In this approach, we first decimated the signal down to a 500 Hz sampling rate. A 3$^{rd}$ order Butterworth highpass filter is then used to filter the signal, using the lower frequency from Table 5 to set the start of the pass frequency. This is done in MATLAB using the *butter* and *filtfilt* functions. Similarly, a lowpass filter is then applied also using a 3$^{rd}$ order Butterworth filter, but with the higher frequency from Table 5.

### *C.2.iii Multi-Taper*

Fourier transforms produce a biased estimate of the spectral content. To alleviate this, the multi-taper method (Thomson, 1982) uses multiple estimates of the spectral content and then combines them back together to reduce the estimation bias. This is done through the use of multiple tapers, which are orthogonal to one another. The average of the collection of orthogonal tapers produces an estimate that has a reduced bias when compared to a Fourier transform estimate (Babadi and Brown, 2014).

We used Slepian tapers (Slepian, 1978) using the Chronux toolbox (Bokil et al., 2010) in our implementation. The window length ($T$) was set to 200 ms, with a step size of 10 ms (for broadband and beta power bands) or 50 ms (for high gamma). The passband ($W$) is set using the values from Table 5. The tapers are computed using the time-bandwidth product ($TW$), which is computed from the window size and the passband bandwidth. The number of tapers is computed as $K = 2*T*W - 1$.

### *C.2.iv Wavelet*

Wavelets (Mallat, 2009) were another alternative explored. The methods described above assume that the data is stationary. Neural processing during speech is dynamic in time, frequency, and space. Thus, a method that can describe the signals in time and frequency simultaneously and properly handle variability is desired. A wavelet-based method has these features and has been shown to yield more robust results in other studies (Jobert et al., 1994; Senhadji and Wendling, 2002; Rosso et al., 2003; Glassman, 2005; Klein et al., 2006), including the relative wavelet power (RWP) which has been shown to associate with the different frequency bands in ECoG (Formaggio et al., 2013).

We explored filtering frequency band results using the Daubechies 4 (db4) wavelet family of functions (Samar et al., 1999). This creates a family of orthogonal functions through successive low-pass and high-pass filtering using the Mallat algorithm (Mallat, 2009).

#### *C.2.v Autoregressive Model*

Others have also used an autoregressive (AR) model to perform filtering (Sturm et al., 2014). We avoid using this method since it builds a linear representation of the ECoG signal as a random process, and thus requires many parameters to fully capture the temporal dependencies. To limit the number of parameters, an AR model could be constructed to only use the previous time sample to predict the next sample. However, this likely does not capture the neural dynamics of speech which have varying time degrees of activation over the span of hundreds of milliseconds.

### C.3 Significant Electrodes and Bad Channel Rejection

Many different methods exist for selecting electrodes that show significant task activity and electrodes that should be rejected due to bad channel characteristics. Here we will briefly review some of the alternatives that we explored.

Many of the techniques used in ECoG analysis have been borrowed from electroencephalography (EEG), which has been around longer and therefore has more established methods. Standardized methods and processing pipelines have been developed for EEG research, such as the EEG processing pipeline (PREP) (Bigdely-Shamlo et al., 2015). We explored leveraging techniques from PREP, primarily for bad

channel rejection. PREP takes a multi-stage approach to removing noisy channels, with robust methods throughout the early processing stages. The pipeline uses four primary methods for removing bad channels: 1) unreasonable amplitudes, 2) lack of cross-channel correlation, 3) lack of cross-channel predictability, and 4) unreasonable frequency noise. Refer to (Bigdely-Shamlo et al., 2015) for more information. It is also still common practice to perform visual inspection for bad channel rejection, e.g. (Pei et al., 2011b).

Determination of significant electrodes, and significant time-points, is largely done through statistical significant tests. There are several variants that are used. A common method to determine if electrodes are significantly responding to a task is to use a t-test. For example, (Hamilton et al., 2018) used bootstrapped t-tests over one second periods and (Hullett et al., 2016) used a one-way t-test with false discovery rate (FDR) to determine significant electrodes. (Moharramipour et al., 2018) looked at multi-variate AR statistical tests and many other variants exist, each making slightly different assumptions about the underlying data distributions. There is no clear winner for which test is best. We used our Kalman filter method to select electrodes that showed a change in trend, which is based on the assumption that the underlying data is distributed as a Gaussian random process. We compared our results to a bootstrapped t-test to validate and found comparable performance between the methods.

### C.4 Epoching

Two alignment cases were discussed within this work, alignment to stimulus presentation and alignment to voicing onset. Both of these alignments are to the first stimulus presented in a trial, with a common baseline for both cases being the time period

from one second to half a second prior to stimulus presentation. The tests that the subjects performed, however, was a word repetition test and there was a paired second stimulus presented in each trial. We also explored additional alignment cases where we aligned to the second stimulus presentation and the subsequent voicing onset.

The baseline period was initially set to the time one second to half second before the second stimulus for alignments in the second half of the trial. This baseline period was found to still have activity present from the response to the first stimulus, which was 3-5 seconds before the second stimulus. Electrode activity had not always returned back to baseline levels during this baseline period, thus making this an invalid baseline, non-task reference. The baseline period from the first stimulus was therefore also used as the baseline for alignments cases in the second half of the trial, as the first baseline period was considered to be the nearest non-task period. The activity that extended into the second baseline period was also found to have an impact during the stimulus presentation period for the second stimulus. These artifacts confounded the findings, making it hard to tease out activity unique to the second stimulus from those remaining or influenced from the first half of the trial. Any results from alignments in the second half of the trial were therefore not used.

We also looked at a longer duration trial that was aligned to the first stimulus presentation and covered through the second vocal response. This was not found to produce meaningful results as there was too much variability in subject responses latencies and the duration between the first and second stimulus was randomly selected from trial to trial.

## C.5 Single Trial

All results are presented from averaged activity across the electrodes within the epoch alignment cases, i.e. ERSPs. Single trial time courses were also explored to see if individual, raw time courses showed the same activity as the characteristic patterns found across the average temporal profiles. Due to the lack of preprocessing, there was a higher degree of variability in the individual time courses. Many individual trials did align well with the characteristic activity patterns during the significant periods, but the higher variability made it more difficult to align across the duration of the trial as the signal to noise ratio is much lower in individual trials versus the epoch. Others in the past have found good single trial alignment, but doing so required additional preprocessing to smooth the signals, e.g. (Conant et al., 2018). We choose not to do that in our study.

# APPENDIX D: Additional Results

This appendix provides additional results that were found using the methodology described in CHAPTER III (§III.3). We present the results with little description or discussion for both completeness and to aid those who venture to further research in this area.

The appendix focusses on three sets of additional results. In §D.1 results for clusters that were only present in a single subject are provided. In §D.2 results from analysis that only uses surface electrodes are presented and contrasted with the results from CHAPTER III, which included depth electrodes. Finally, §D.3 presents results when high gamma power activity is switched with beta band power, as discussed in §C.2.

## D.1 Single Subject Clusters

CHAPTER III only presented results for clusters that were present in at least two subjects, with all resulting clusters coming from at least three subjects. Here we provide results for the electrodes that were not included in that analysis and formed clusters only present within a single subject. Several of the clusters contained only a single electrode.

Single subject clusters were seen in both the stimulus presentation alignment condition and when aligned to voicing onset. Results for the two conditions are presented in §D.1.i and §D.1.ii, respectively. Results are further broken down by the number of electrodes that fall within each cluster.

### *D.1.i Stimulus Presentation Alignment Single Subject Clusters*

Results for single subject clusters in the stimulus presentation alignment case are broken out across the following three figures. Figure 32 shows results for the cluster with the most electrodes, which displays high gamma suppression during the task period. We refer to this cluster as high gamma suppression. Suppression of high gamma activity during self-generated speech was found previously within the auditory cortex, with variance across subjects (Flinker et al., 2010). Figure 33 shows three clusters that had a limited number of electrodes. Figure 34 shows the final three clusters that only had one electrode each.

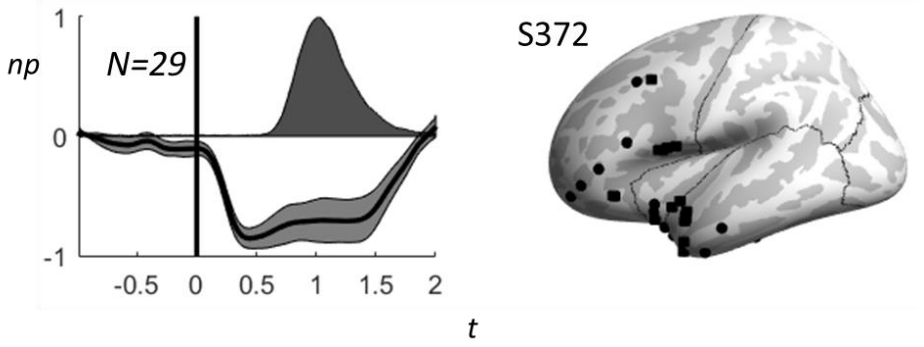


**Figure 32: Stimulus Alignment Single Subject High Gamma Suppression Cluster**

Left: Normalized power (np) over time (t) for 29 electrodes that are clustered together within Subject 372. Solid vertical line indicates stimulus presentation timing. Average audio signal amplitude indicated by gray shaded region. Right: Subject 372 active electrodes for cluster.

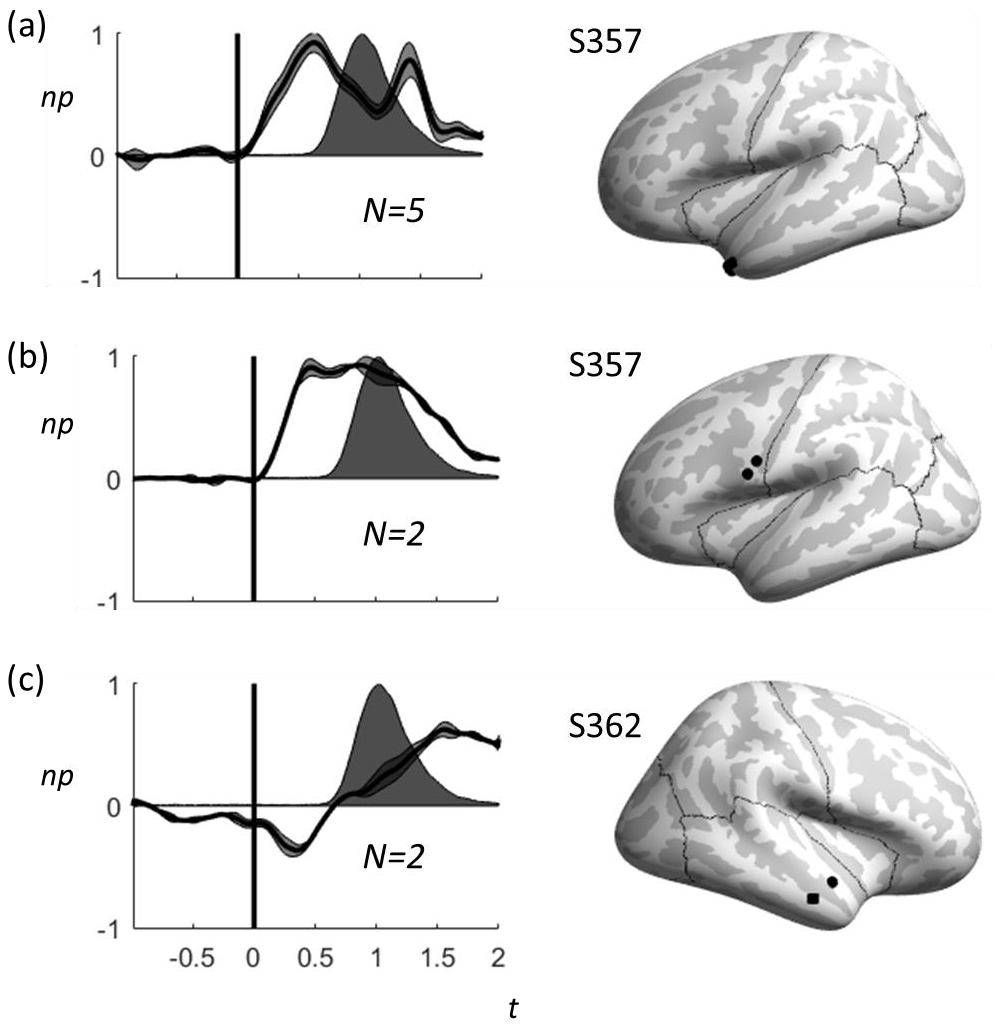


**Figure 33: Stimulus Alignment Single Subject Clusters with more than 1 Electrode**

Additional single subject clusters for stimulus alignment case with more than one electrode. (a,b) Subject 357 clusters with 5 and 2 electrodes, respectively. (c) Subject 362 cluster with 2 electrodes. Left: Normalized power (np) over time (t). Solid vertical lines indicate stimulus presentation timing. Average audio signal amplitude indicated by gray shaded region. Right: Subject active electrodes for clusters.

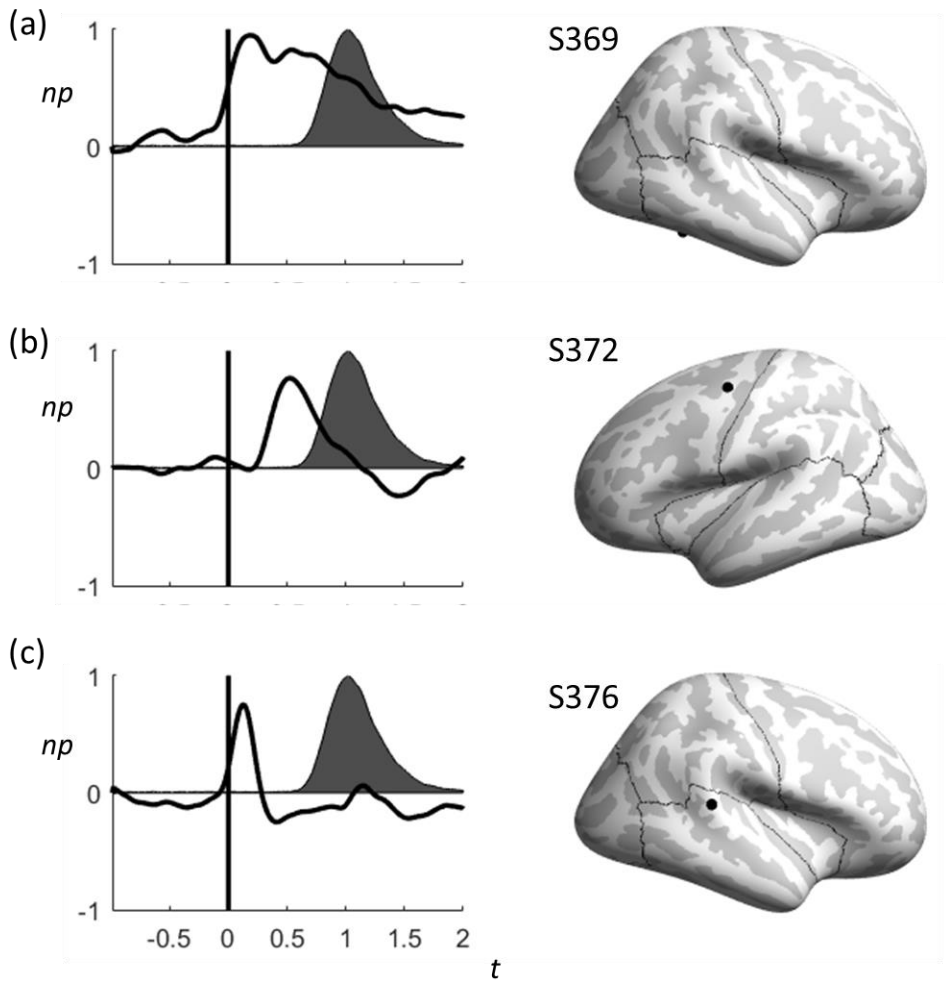


**Figure 34: Stimulus Alignment Single Subject Clusters with Only 1 Electrode**

Single subject clusters that only have one electrode. (a) Subject 369, (b) Subject 372, and (c) Subject 376. Left: Normalized power (np) over time (t). Solid vertical lines indicate stimulus presentation timing. Average audio signal amplitude indicated by gray shaded region. Right: Subject active electrode for clusters.

### *D.1.ii Voicing Onset Alignment Single Subject Clusters*

Results for the voicing onset alignment case are broken out across the following three figures. Figure 35 shows results for the cluster with the most electrodes. This cluster matches the high gamma suppression cluster in Figure 32 for the stimulus

alignment case. Figure 36 shows three clusters with limited number of electrodes. Figure 37 shows four clusters that only had one electrode each.

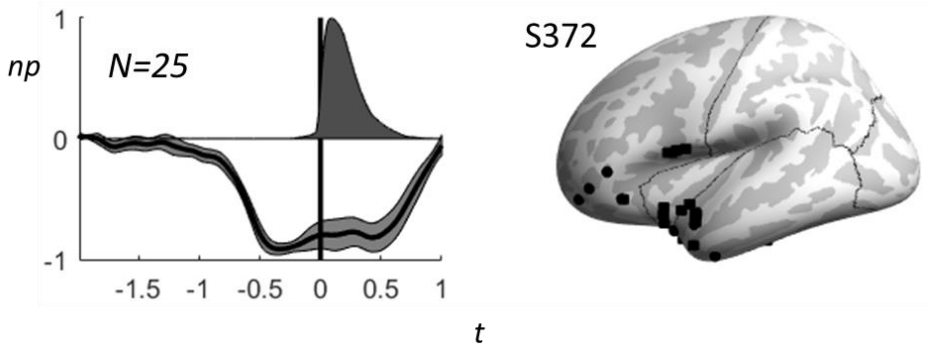


**Figure 35: Voicing Alignment Single Subject High Gamma Suppression Cluster**

Left: Normalized power (np) over time (t) for 25 electrodes that are clustered together within Subject 372. Solid vertical line indicates time of voicing onset. Average audio signal amplitude indicated by gray shaded region. Right: Subject 372 active electrodes for cluster.

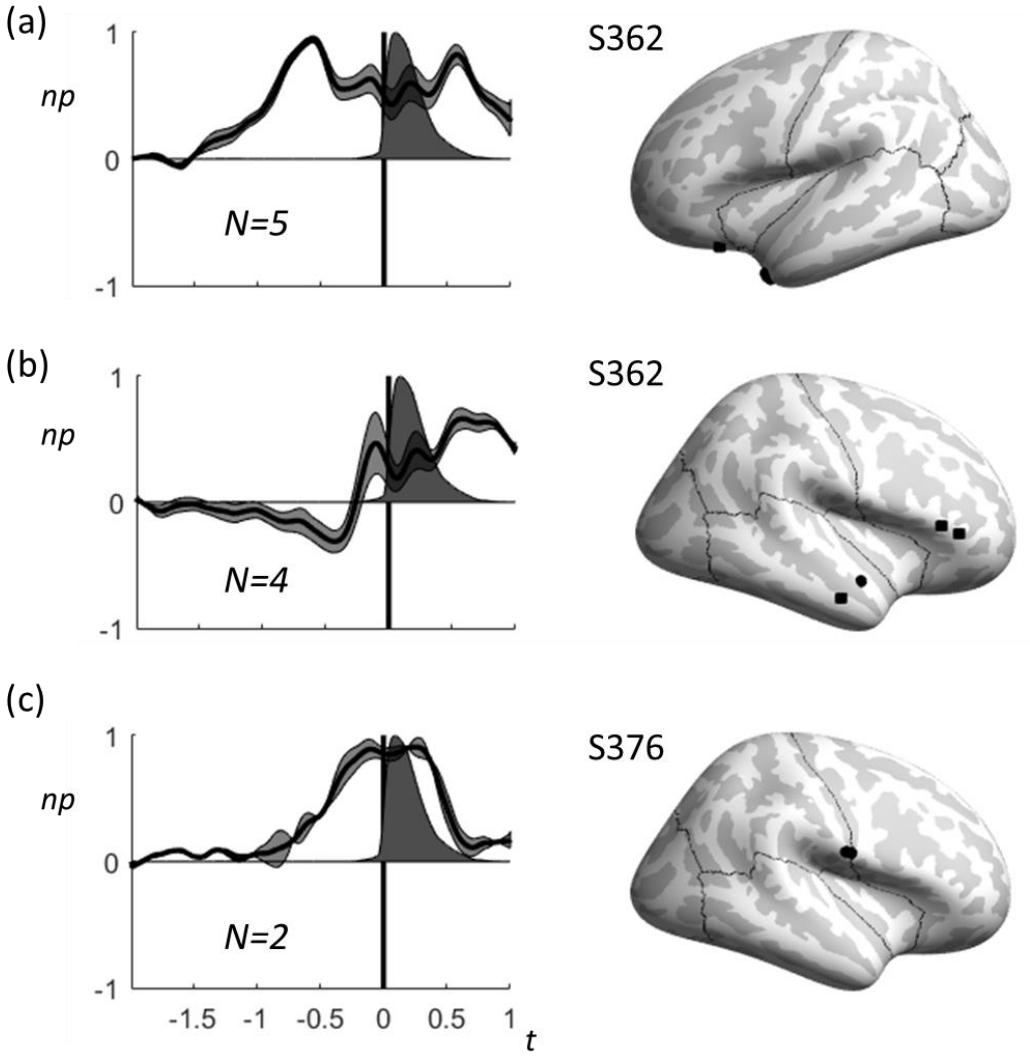


**Figure 36: : Voicing Alignment Single Subject Clusters with more than 1 Electrode**

Additional single subject clusters from voicing onset alignment case with more than one electrode. (a,b) Subject 362 clusters with 5 and 4 electrodes, respectively. (c) Subject 376 cluster with 2 electrodes. Left: Normalized power (np) over time (t). Solid vertical lines indicate time of voicing onset. Average audio signal amplitude indicated by gray shaded region. Right: Subject active electrodes for clusters.

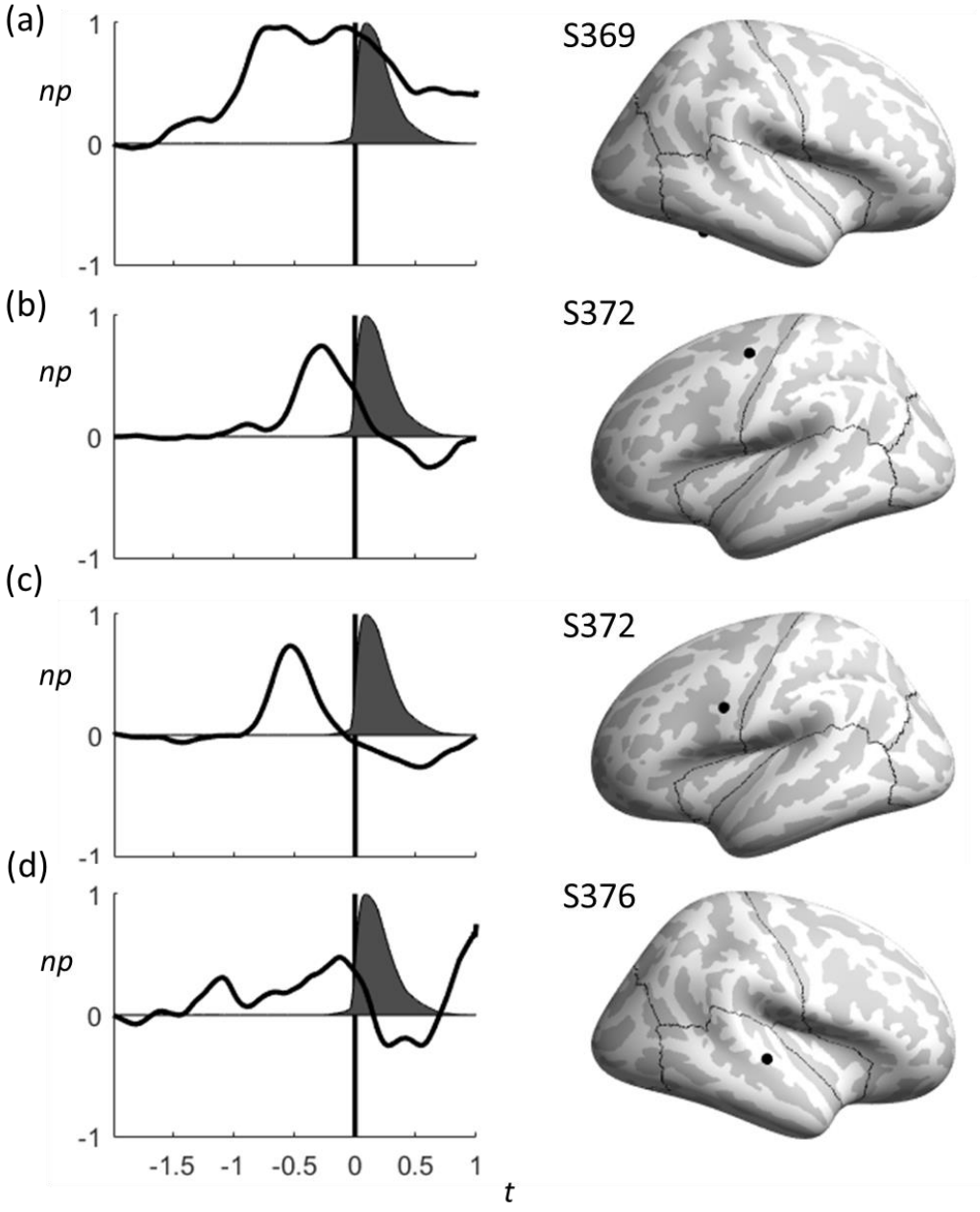


**Figure 37: Voicing Alignment Single Subject Clusters with Only 1 Electrode**

Single subject clusters that only have one electrode. (a) Subject 369, (b,c) Subject 372, and (d) Subject 376. Left: Normalized power (np) over time (t). Solid vertical lines indicate time of voicing onset. Average audio signal amplitude indicated by gray shaded region. Right: Subject active electrode for clusters.

## D.2 Electrocorticography Surface Electrodes Only

The results of CHAPTER III contained both surface and depth electrocorticographic (ECoG) electrodes. In this section we present results if only surface ECoG electrodes are used. Most ECoG studies are constrained to only surface electrodes, due to clinical reasons. Thus, this section provides a direct comparison to those studies. This section also provides evidence that the canonical characteristic time courses found in CHAPTER III still hold when only considering surface electrodes, and thus the results of CHAPTER III can be directly compared to prior studies on their own.

Resulting characteristic time courses are shown in Figure 38 (a) and (b) for stimulus presentation and voicing onset alignments, respectively. Six time courses were found in the stimulus presentation case and given the same names as in CHAPTER III. Seven time courses were found in the voicing onset case. Six of them align with those found in CHAPTER III and are given the same names. One new time course has a very broad plateau of activity and is simply referred to as New.

Individual cluster results, including location of electrodes, are presented in Figure 39 and Figure 40 for the stimulus presentation and voicing onset alignment cases, respectively. Figure 41 shows very high degree of temporal profile similarity between clusters found using only surface electrodes (dashed colored lines) and clusters found using all electrodes, i.e. CHAPTER III (solid colored lines). All figures maintain colors with those used in CHAPTER III.

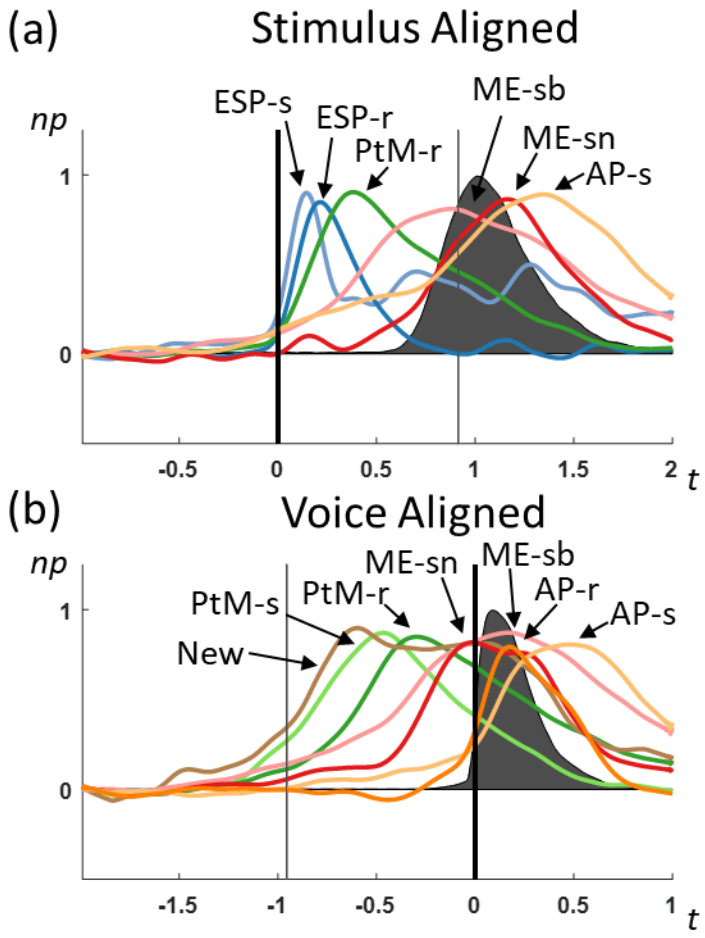


**Figure 38: Canonical Activity Patterns for Surface Electrodes Only**

Characteristic high gamma power (normalized power (np)) time courses (time (t) in seconds) for (a) stimulus presentation and (b) voicing onset alignment conditions when only surface electrodes are present in the analysis. Average audio signal amplitude indicated by gray shaded region. (a) Six characteristic time courses from stimulus presentation aligned case: Early Stimulus Processing – symmetric (ESP-s), Early Stimulus Processing – ramp (ESP-r), Phonological-to-Motor – ramp (PtM-r), Motor Execution – symmetric broad (ME-sb), Motor Execution – symmetric narrow (ME-sn), and Auditory Processing – symmetric (AP-s). Stimulus presentation occurred at t=0, shown with vertical solid black line. Average voicing onset per cluster shown in fainter vertical black lines (different lines since not all subjects showed a response for all clusters). (b) *Seven* characteristic time courses from voicing onset aligned cases: Phonological-to-Motor – symmetric (PtM-s), Phonological-to-Motor – ramp (PtM-r), Motor Execution – symmetric narrow (ME-sn), Motor Execution – symmetric broad (ME-sb), Auditory Processing – ramp (AP-r), Auditory Processing – symmetric (AP-s), and a new time course (New) not seen in the analysis with depth electrodes included. Time axis reference to voicing onset (t=0) with a solid vertical line for the time of voicing onset. Fainter vertical lines for average stimulus presentation time per cluster. Colors maintained with those used in CHAPTER III.

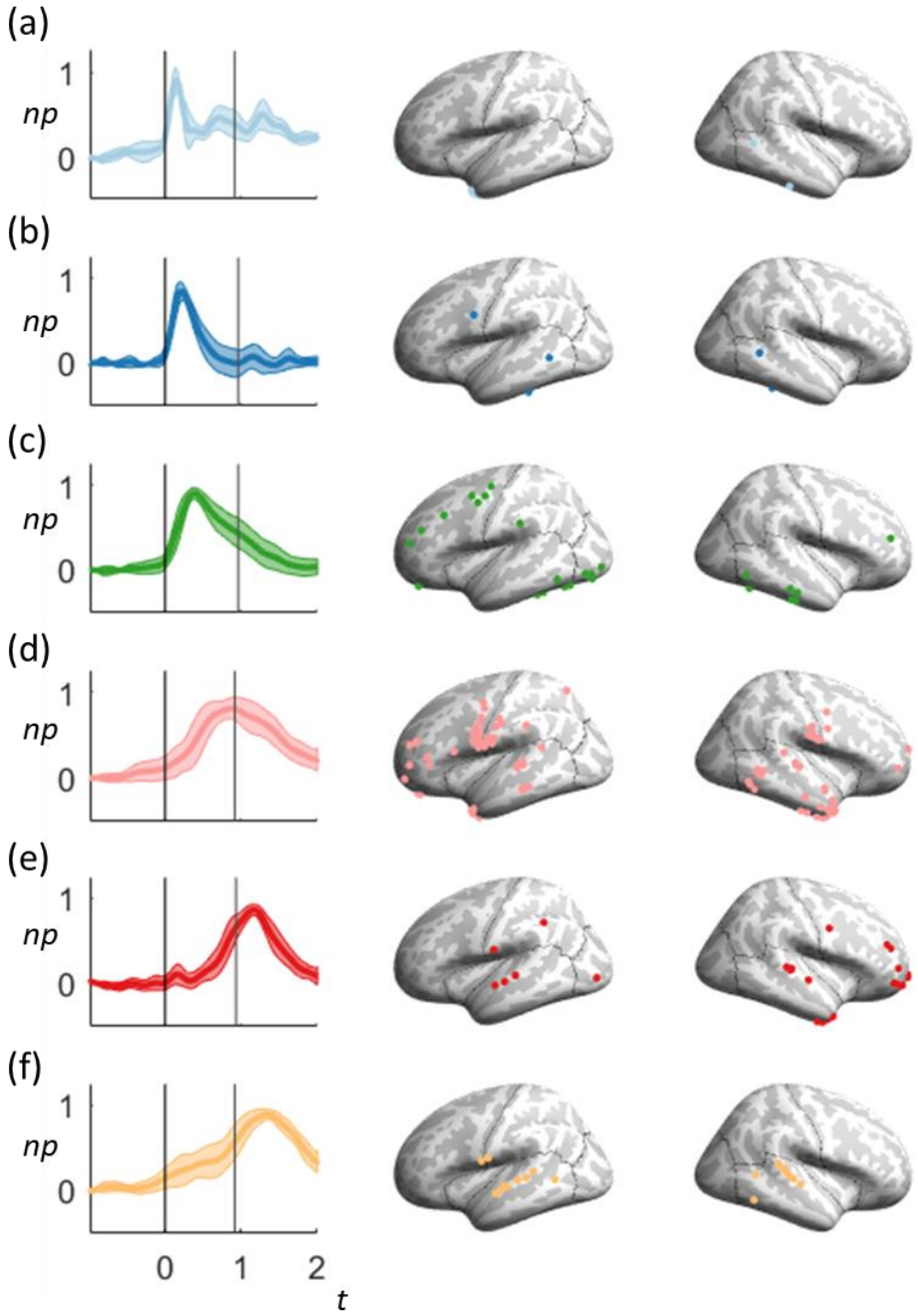


**Figure 39: Surface Electrode Clusters for Stimulus Presentation Alignment**

High gamma cluster activity patterns are shown on the left (normalized power (np) over time (t), in seconds), electrode locations for clusters on the right. Solid vertical lines indicate stimulus presentation (t=0), and fainter vertical lines show average location of voicing onset. (a) Early Stimulus Processing – symmetric (ESP-s), (b) Early Stimulus Processing – ramp (ESP-r), (c) Phonological-to-Motor processing – ramp (PtM-r), (d) Motor Execution – symmetric broad (ME-sb), (e) Motor Execution – symmetric narrow (ME-sn), and (f) Auditory Processing – symmetric (AP-s).

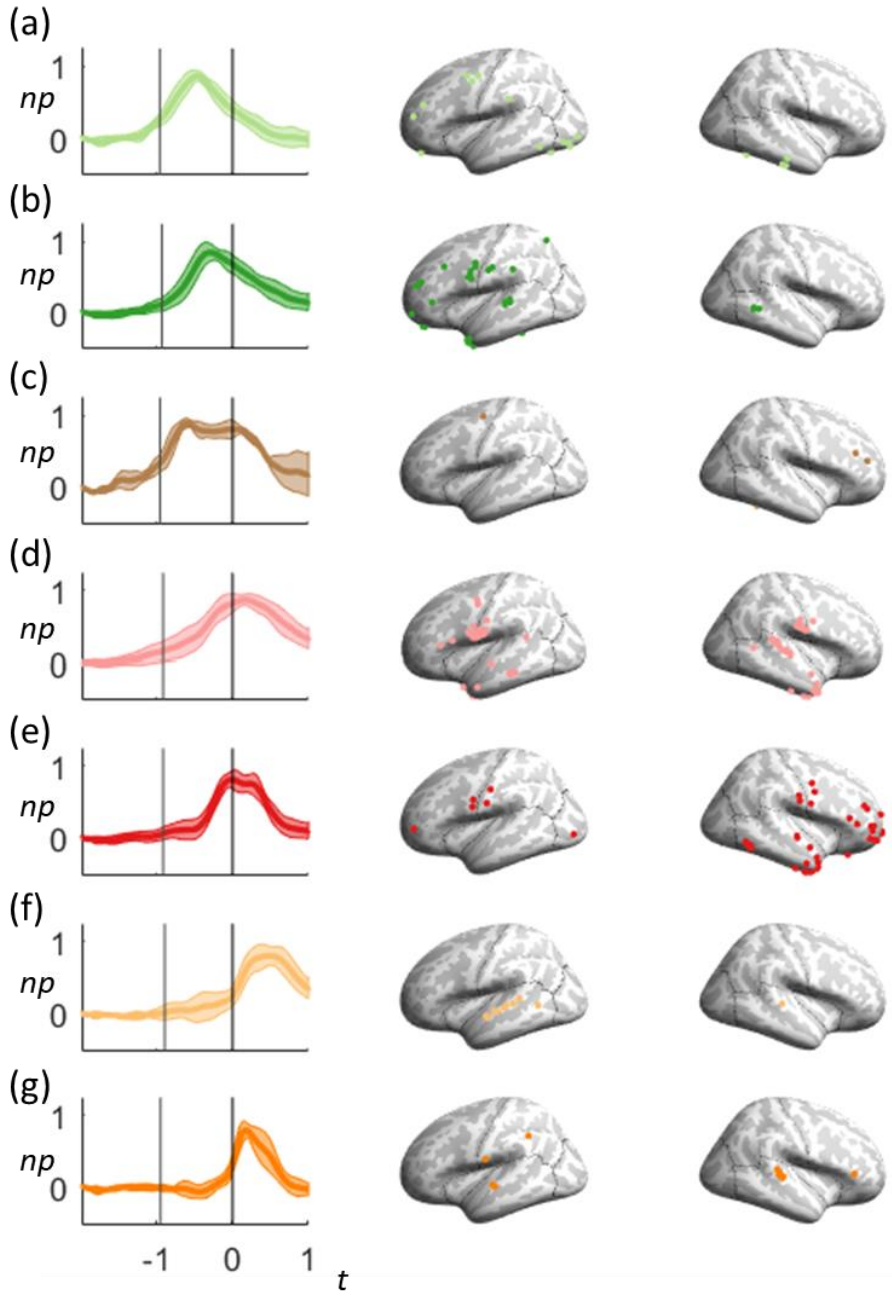


**Figure 40: : Surface Electrode Clusters for Voicing Onset Alignment**

High gamma cluster activity patterns are shown on the left (normalized power (np) over time (t), in seconds), electrode locations for clusters on the right. Solid vertical lines indicate voicing onset (t=0), and fainter vertical lines show average location of stimulus presentation. (a) Phonological-to-Motor processing – symmetric (PtM-s), (b) Phonological-to-Motor processing – ramp (PtM-r), (c) previously unseen cluster (New) (d) Motor Execution – symmetric broad (ME-sb), (e) Motor Execution – symmetric narrow (ME-sn), (f) Auditory Processing – symmetric (AP-s), and (g) Auditory Processing – ramp (AP-r).

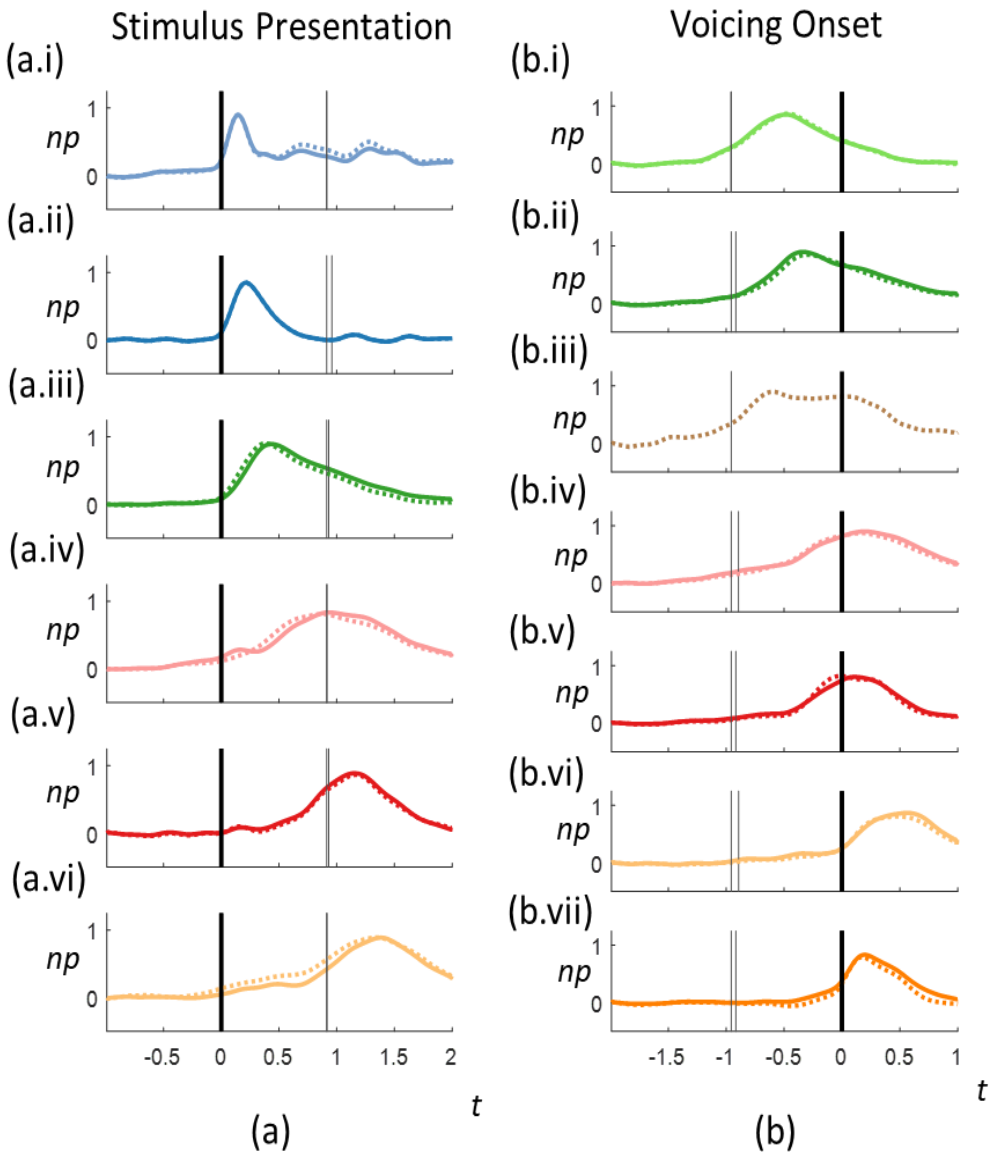


**Figure 41: Cluster Comparison of Surface Electrodes to All Electrodes**

Characteristic high gamma power (normalized power (np)) over time (t) for clusters from (a) stimulus presentation and (b) voicing onset alignment cases. Solid vertical lines indicate alignment condition in which the cluster was identified, and fainter vertical lines show average location of other alignment condition. Solid color lines are for characteristic activity from clustering with all electrodes, dashed lines for surface electrodes only. (a.i) Early Stimulus Processing – symmetric (ESP-s), (a.ii) Early Stimulus Processing – ramp (ESP-r), (a.iii) Phonological-to-Motor processing – ramp (PtM-r), (a.iv) Motor Execution – symmetric broad (ME-sb), (a.v) Motor Execution – symmetric narrow (ME-sn), and (a.vi) Auditory Processing – symmetric (AP-s). (b.i) Phonological-to-Motor processing – symmetric (PtM-s), (b.ii) PtM-r, (b.iii) previously unseen cluster (New) (b.iv) ME-sb, (b.v) ME-sn, (b.vi) AP-s, and (b.vii) Auditory Processing – ramp (AP-r). Colors maintained with those used in CHAPTER III.

## D.3 Beta Frequency Band

Beta frequency rhythms have been found during task-related activity (Wang, 2010) and motor preparation (Rubino, 2006), including desynchronization, or beta suppression (Pfurtscheller and Lopes da Silva, 1999). The primary focus of our work has been on high gamma power. Here, we provide clustering results from the beta frequency band to demonstrate how the methodology translates to other frequency bands of interest and has the potential to be more widely used.

The characteristic *beta* power time courses are shown in Figure 42 for the stimulus presentation alignment case. Figure 43 further breaks down these findings by cluster and shows the electrode locations. In Figure 44, the characteristic beta power time courses are shown for the voicing onset alignment case. Figure 45 presents the details of each individual cluster with the electrode locations for the voicing onset alignment case.

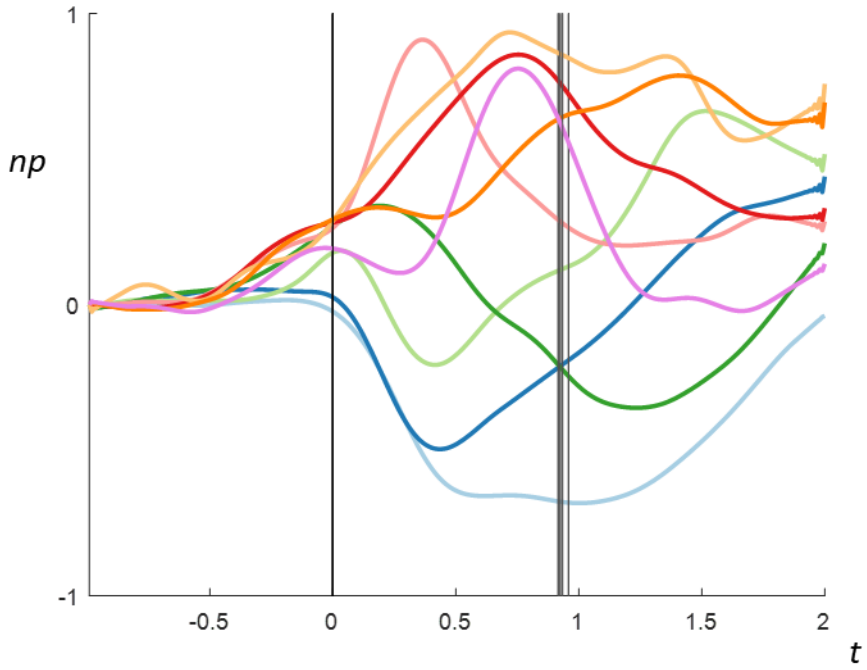


**Figure 42: Beta Power Characteristic Time Courses for Stimulus Presentation Alignment**

Characteristic time courses (normalized power (np)) over time (t). Solid vertical lines indicate stimulus presentation (t=0), and fainter vertical lines show average location of voicing onset (different lines since not all subjects showed a response for all clusters). Different colors used to display each cluster.

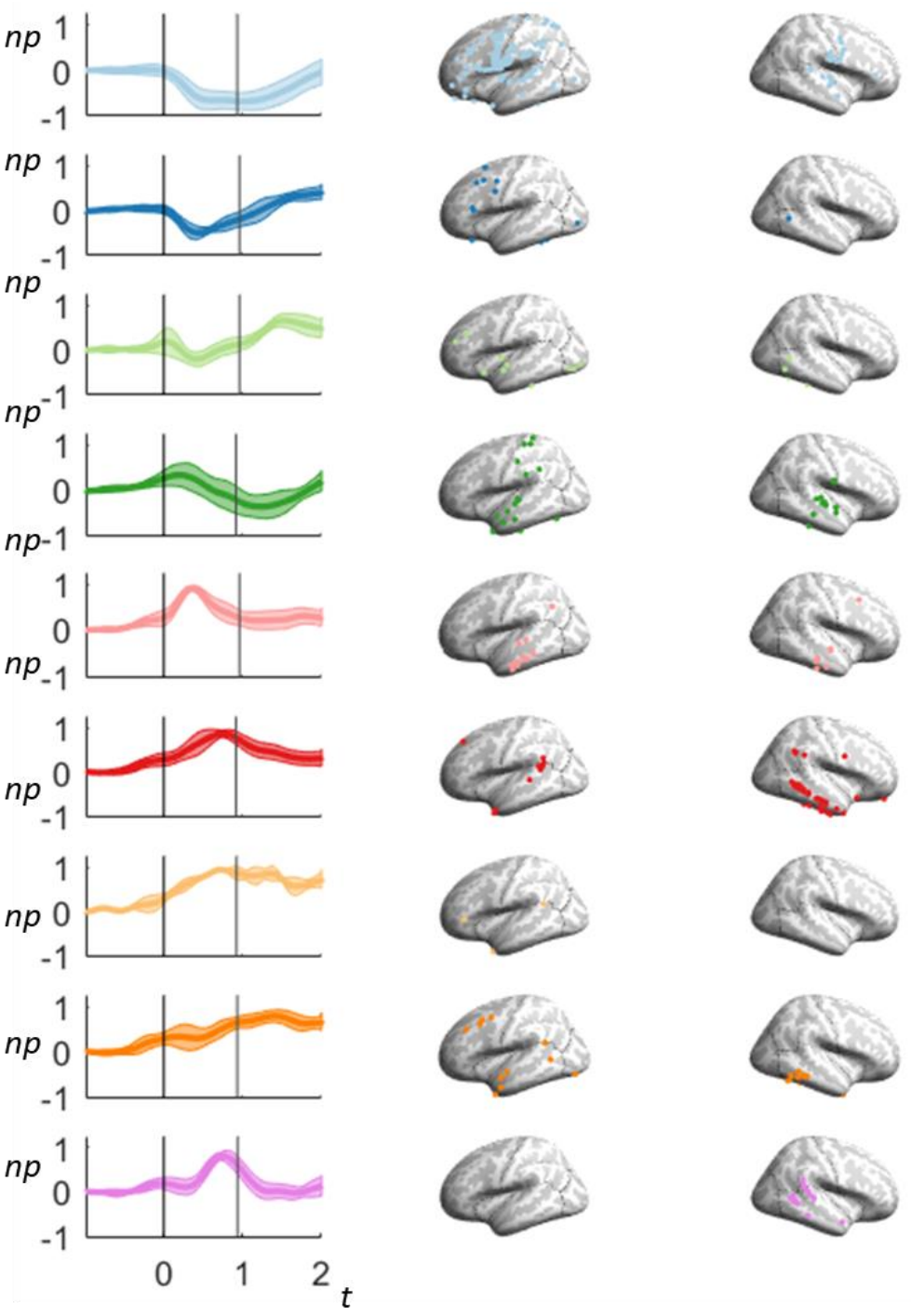


**Figure 43: Individual Beta Power Clusters for Stimulus Presentation Alignment**

Left: Characteristic time courses (normalized power (np)) over time (t). Solid vertical lines indicate stimulus presentation (t=0), and fainter vertical lines show average location of voicing onset. Right: Electrode locations for clusters. Each row displays a different cluster, each in a different color.

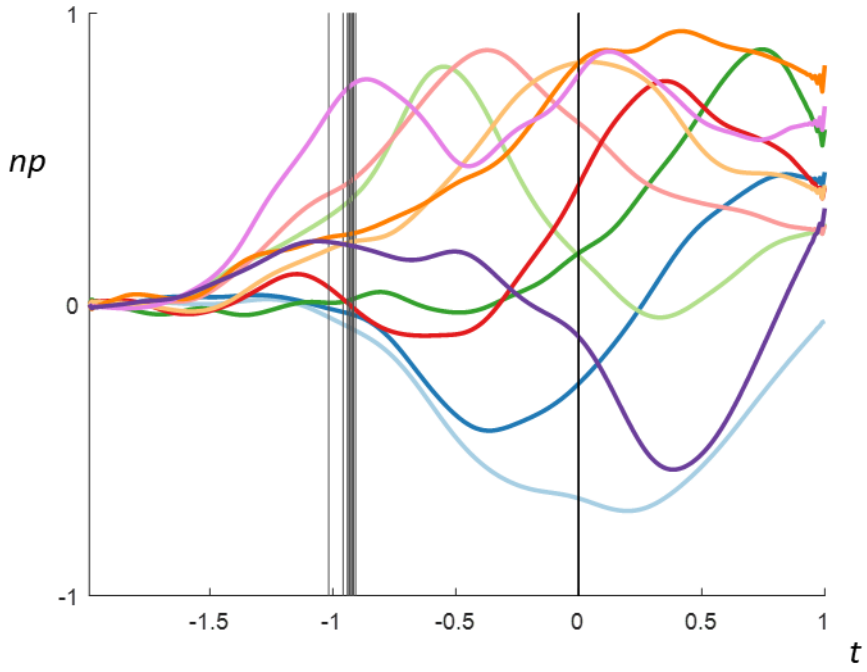


**Figure 44: Beta Power Characteristic Time Courses for Voicing Onset Alignment**

Characteristic time courses (normalized power (np)) over time (t). Solid vertical lines indicate voicing onset (t=0), and fainter vertical lines show average location of stimulus presentation (different lines since not all subjects showed a response for all clusters). Different colors used to display each cluster.

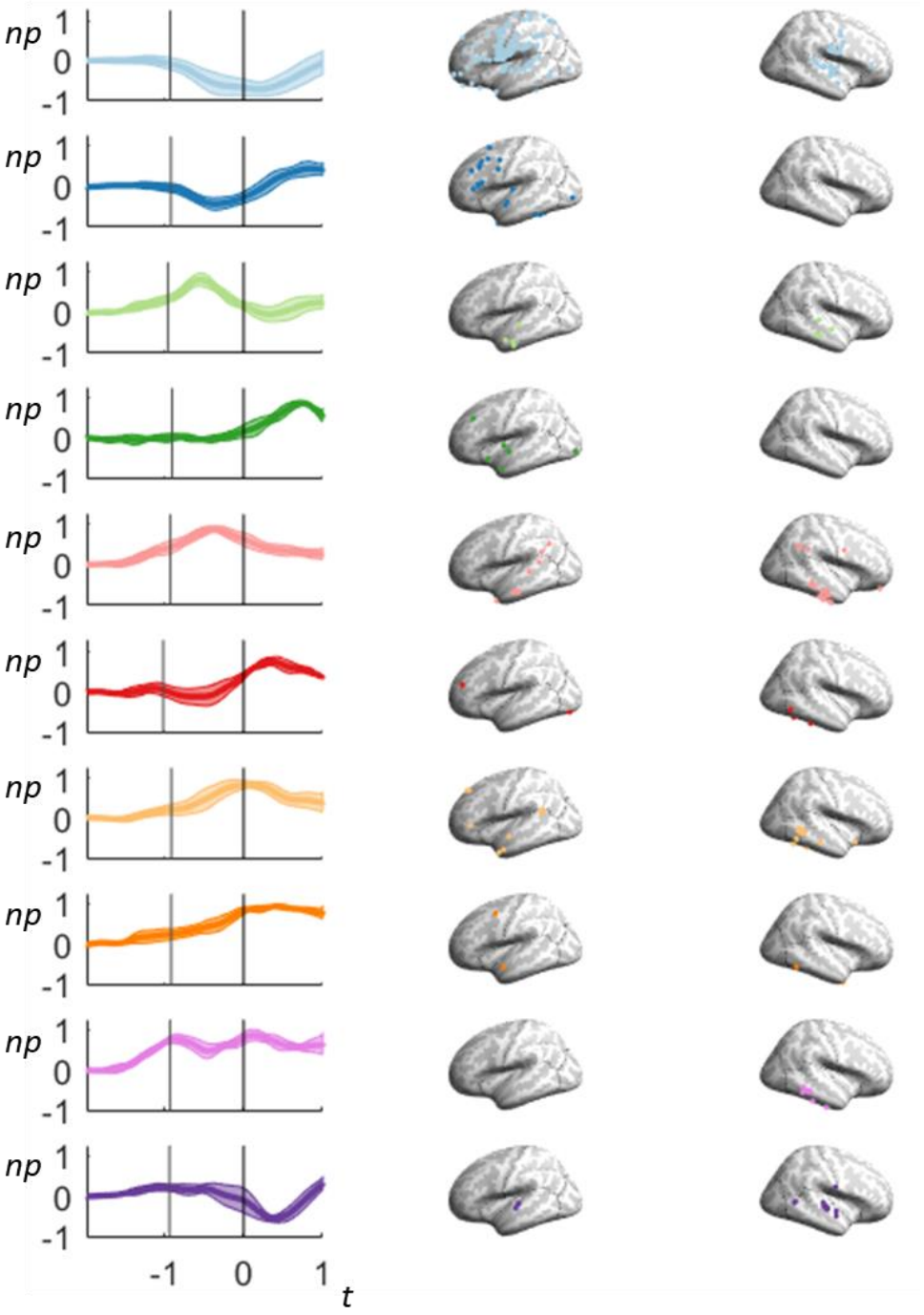


**Figure 45: Individual Beta Power Clusters for Voicing Onset Alignment**

Left: Characteristic time courses (normalized power (np)) over time (t). Solid vertical lines indicate voicing onset (t=0), and fainter vertical lines show average location of stimulus presentation. Right: Electrode locations for clusters. Each row displays a different cluster, each in a different color.

## BIBLIOGRAPHY

Adams RP, MacKay DJC (2007) Bayesian Online Changepoint Detection. arXiv e-prints arXiv:0710.3742 [stat].

Aghabozorgi S, Seyed Shirkhorshidi A, Ying Wah T (2015) Time-series clustering – A decade review. Information Systems 53:16–38.

Ahveninen J, Jääskeläinen IP, Raij T, Bonmassar G, Devore S, Hämäläinen M, Levänen S, Lin F-H, Sams M, Shinn-Cunningham BG, Witzel T, Belliveau JW (2006) Task-modulated "what" and "where" pathways in human auditory cortex. Proceedings of the National Academy of Sciences of the United States of America 103:14608–14613.

Altmann CF, Gomes de Oliveira Júnior C, Heinemann L, Kaiser J (2010) Processing of spectral and amplitude envelope of animal vocalizations in the human auditory cortex. Neuropsychologia 48:2824–2832.

Aminikhanghahi S, Cook DJ (2017) A Survey of Methods for Time Series Change Point Detection. Knowledge Information Systems 51:339–367.

Anderson JM, Gilmore R, Roper S, Crosson B, Bauer RM, Nadeau S, Beversdorf DQ, Cibula J, Rogish M, Kortencamp S, Hughes JD, Gonzalez Rothi LJ, Heilman KM (1999) Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. Brain and Language 70:1–12.

Angrick M, Herff C, Mugler E, Tate MC, Slutzky MW, Krusienski DJ, Schultz T (2019) Speech synthesis from ECoG using densely connected 3D convolutional neural networks. Journal of Neural Engineering 16:036019.

Anton H (2010) Elementary Linear Algebra, 10 edition. Hoboken, NJ: Wiley.

Anumanchipalli GK, Chartier J, Chang EF (2019) Speech synthesis from neural decoding of spoken sentences. Nature 568:493–498.

Arya R, Wilson JA, Vannest J, Byars AW, Greiner HM, Buroker J, Fujiwara H, Mangano FT, Holland KD, Horn PS, Crone NE, Rose DF (2015) Electrocorticographic language mapping in children by high-gamma synchronization during spontaneous conversation: comparison with conventional electrical cortical stimulation. Epilepsy Research 110:78–87.

Babadi B, Brown EN (2014) A Review of Multitaper Spectral Analysis. Institute of Electrical and Electronics Engineers Transactions of Biomedical Engineering 61:1555–1564.

Baldo JV, Klostermann EC, Dronkers NF (2008) It's either a cook or a baker: patients with conduction aphasia get the gist but lose the trace. Brain and Language 105:134–140.

Beaulieu NC (1989) A simple series for personal computer computation of the error function Q(.). Institute of Electrical and Electronics Engineers Transactions on Communications 37:989–991.

Berezutskaya J, Freudenburg ZV, Güçlü U, Gerven MAJ van, Ramsey NF (2017) Neural Tuning to Low-Level Features of Speech throughout the Perisylvian Cortex. Journal of Neuroscience 37:7906–7920.

Berndt DJ, Clifford J (1994) Using Dynamic Time Warping to Find Patterns in Time Series. In: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, pp 359–370. Seattle, WA: Association for the Advancement of Artificial Intelligence Press.

Bigdely-Shamlo N, Mullen T, Kothe C, Su K-M, Robbins KA (2015) The PREP pipeline: standardized preprocessing for large-scale EEG analysis. Frontiers of Neuroinformatics 9.

Bilecen D, Scheffler K, Schmid N, Tschopp K, Seelig J (1998) Tonotopic organization of the human auditory cortex as detected by BOLD-FMRI. Hearing Research 126:19–27.

Blakely T, Miller KJ, Rao RPN, Holmes MD, Ojemann JG (2008) Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids. Conference Proceedings of Institute of Electrical and Electronics Engineers Engineering in Medicine and Biology Society 2008:4964–4967.

Bohland JW, Guenther FH (2006) An fMRI investigation of syllable sequence production. Neuroimage 32:821–841.

Bokil H, Andrews P, Kulkarni JE, Mehta S, Mitra P (2010) Chronux: A Platform for Analyzing Neural Signals. Journal of Neuroscience Methods 192:146–151.

Borjesson P, Sundberg C- (1979) Simple Approximations of the Error Function Q(x) for Communications Applications. Institute of Electrical and Electronics Engineers Transactions on Communications 27:639–643.

Bouchard KE, Chang EF (2014) Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. Journal of Neuroscience 34:12662–12677.

Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional Organization of Human Sensorimotor Cortex for Speech Articulation. Nature 495:327–332.

Brennan J, Pylkkänen L (2012) The time-course and spatial distribution of brain activity associated with sentence processing. NeuroImage 60:1139–1148.

Broca P (1861) Perte de la parole, ramollissement chronique et destruction partielle du lobe anterieur gauche du cerveau. Bulletin de la Société d'Anthropologie 2:235–238.

Broca P (1865) Sur le siège de la faculté du langage articulé. Bulletins de la Société d'Anthropologie 6:377–393.

Brumberg JS, Castro N, Rao A (2015) Temporal dynamics of the speech readiness potential, and its use in a neural decoder of speech-motor intention. In: Annual Conference of the International Speech Communication Association, pp 1126–1130.

Brumberg JS, Krusienski DJ, Chakrabarti S, Gunduz A, Brunner P, Ritaccio AL, Schalk G (2016) Spatio-Temporal Progression of Cortical Activity Related to Continuous Overt and Covert Speech Production in a Reading Task. Public Library of Science One 11:e0166872.

Brumberg JS, Nieto-Castanon A, Kennedy PR, Guenther FH (2010) Brain–computer interfaces for speech communication. Speech Communication 52:367–379.

Brumberg JS, Wright EJ, Andreasen DS, Guenther FH, Kennedy PR (2011) Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech motor cortex. Frontiers of Neuroscience 5.

Brunner P, Dijkstra K, Coon WG, Mellinger J, Ritaccio AL, Schalk G (2017) An ECoG-Based BCI Based on Auditory Attention to Natural Speech. In: Brain-Computer Interface Research, pp 7–19 Springer Briefs in Electrical and Computer Engineering. Springer.

Buchsbaum BR, Hickok G, Humphries C (2001) Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. Cognitive Science 25:663–678.

Buchsbaum BR, Olsen RK, Koch P, Berman KF (2005) Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. Neuron 48:687–697.

Buchweitz A, Mason RA, Tomitch LMB, Just MA (2009) Brain activation for reading and listening comprehension: An fMRI study of modality effects and individual differences in language comprehension. Psychology & Neuroscience 2:111–123.

Butterworth S (1930) On the Theory of Filter Amplifiers. Wireless Engineer:536–541.

Canolty RT, Edwards E, Dalal SS, Soltani M, Nagarajan SS, Kirsch HE, Berger MS, Barbaro NM, Knight RT (2006) High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex. Science 313:1626–1628.

Chang EF, Edwards E, Nagarajan SS, Fogelson N, Dalal SS, Canolty RT, Kirsch HE, Barbaro NM, Knight RT (2011) Cortical Spatio-temporal Dynamics Underlying Phonological Target Detection in Humans. Journal of Cognitive Neuroscience 23:1437–1446.

Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF (2013) Human cortical sensorimotor network underlying feedback control of vocal pitch. Proceedings of the National Academy of Sciences of the United States of America 110:2653–2658.

Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT (2010) Categorical Speech Representation in Human Superior Temporal Gyrus. Nature Neuroscience 13:1428–1432.

Chaumon M, Kveraga K, Barrett LF, Bar M (2014) Visual Predictions in the Orbitofrontal Cortex Rely on Associative Content. Cerebral Cortex 24:2899–2907.

Cheney PD, Fetz EE (1980) Functional classes of primate corticomotoneuronal cells and their relation to active force. Journal of Neurophysiology 44:773–791.

Cheung C, Hamiton LS, Johnson K, Chang EF (2016) The auditory representation of speech sounds in human motor cortex. Elife 5.

Chintaluri C, Wójcik DK (2015) A novel method for spatial source localization using ECoG and SEEG recordings in human epilepsy patients. Biomedicine Central Neuroscience 16:P286.

Cogan GB, Thesen T, Carlson C, Doyle W, Devinsky O, Pesaran B (2014) Sensory-motor transformations for speech occur bilaterally. Nature 507:94–98.

Collard MJ, Fifer MS, Benz HL, McMullen DP, Wang Y, Milsap GW, Korzeniewska A, Crone NE (2016) Cortical subnetwork dynamics during human language tasks. Neuroimage 135:261–272.

Conant DF, Bouchard KE, Leonard MK, Chang EF (2018) Human Sensorimotor Cortex Control of Directly Measured Vocal Tract Movements during Vowel Production. Journal of Neuroscience 38:2955–2966.

Cooley JW, Tukey JW (1965) An Algorithm for the Machine Calculation of Complex Fourier Series. Mathematics of Computation 19:297–301.

Craig JW (1991) A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations. In: Military Communications 91 - Conference record, pp 571–575. McLean, VA: Institute of Electrical and Electronics Engineers.

Crone NE, Boatman D, Gordon B, Hao L (2001a) Induced electrocorticographic gamma activity during auditory perception. Clinical Neurophysiology 112:565–582.

Crone NE, Hao L, Hart J, Boatman D, Lesser RP, Irizarry R, Gordon B (2001b) Electrocorticographic gamma activity during word production in spoken and sign language. Neurology 57:2045–2053.

Crone NE, Sinai A, Korzeniewska A (2006) High-frequency gamma oscillations and human brain mapping with electrocorticography. Progress in Brain Research 159:275–295.

Dell GS (1986) A spreading-activation theory of retrieval in sentence production. Psychology Review 93:283–321.

Dichter BK, Bouchard KE, Chang EF (2016) Dynamic Structure of Neural Variability in the Cortical Representation of Speech Sounds. Journal of Neuroscience 36:7453–7463.

Ding C, He X, Simon H (2005) On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering. In: Proceedings of the 2005 Society for Industrial and Applied Mathematics International Conference on Data Mining, pp 606–610 Proceedings. Society for Industrial and Applied Mathematics.

Ding CHQ, Li T, Jordan MI (2010) Convex and Semi-Nonnegative Matrix Factorizations. Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence 32:45–55.

Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical tracking of hierarchical linguistic structures in connected speech. Nature Neuroscience 19:158–164.

Edwards E, Nagarajan SS, Dalal SS, Canolty RT, Kirsch HE, Barbaro NM, Knight RT (2010) Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage 50:291–301.

Edwards E, Soltani M, Kim W, Dalal SS, Nagarajan SS, Berger MS, Knight RT (2009) Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. Journal of Neurophysiology 102:377–386.

Eliseyev A, Aksenova T (2016) Penalized Multi-Way Partial Least Squares for Smooth Trajectory Decoding from Electrocorticographic (ECoG) Recording. Public Library of Science One 11:e0154878.

Fischl B (2012) FreeSurfer. Neuroimage 62:774–781.

Flinker A, Chang EF, Barbaro NM, Berger MS, Knight RT (2011) Sub-centimeter language organization in the human temporal lobe. Brain and Language 117:103–109.

Flinker A, Chang EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT (2010) Single-trial speech suppression of auditory cortex activity in humans. Journal of Neuroscience 30:16643–16650.

Flinker A, Korzeniewska A, Shestyuk AY, Franaszczuk PJ, Dronkers NF, Knight RT, Crone NE (2015) Redefining the role of Broca's area in speech. Proceedings of the National Academy of Science of the United States of America 112:2871–2875.

Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud A-L (2014) The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nature Communications 5:4694.

Formaggio E, Storti SF, Tramontano V, Casarin A, Bertoldo A, Fiaschi A, Talacchi A, Sala F, Toffolo GM, Manganotti P (2013) Frequency and time-frequency analysis of intraoperative ECoG during awake brain stimulation. Frontiers of Neuroengineering 6.

Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron 40:859–869.

Frey BJ, Dueck D (2007) Clustering by Passing Messages Between Data Points. Science 315:972–976.

Friederici AD (2012) The cortical language circuit: from auditory perception to sentence comprehension. Trends in Cognitive Science 16:262–268.

Friederici AD (2017) Language in Our Brain: The Origins of a Uniquely Human Capacity. Cambridge, MA: MIT Press.

Garagnani M, Pulvermüller F (2013) Neuronal correlates of decisions to speak and act: Spontaneous emergence and dynamic topographies in a computational model of frontal and temporal areas. Brain and Language 127:75–85.

Gasser T, Bächer P, Möcks J (1982) Transformations towards the normal distribution of broad band spectral parameters of the EEG. Electroencephalography and Clinical Neurophysiology 53:119–124.

Geffner H (2018) Model-free, Model-based, and General Intelligence. arXiv e-prints arXiv:1806.02308 [cs].

Geschwind N (1965) Disconnexion syndromes in animals and man. I. Brain 88:237–294.

Geschwind N (1979) Specializations of the human brain. Scientific American 241:180–199.

Ghosh SS, Tourville JA, Guenther FH (2008) A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. Journal of Speech Language and Hearing Research 51:1183–1202.

Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A (2000) Representation of the temporal envelope of sounds in the human brain. Journal of Neurophysiology 84:1588–1598.

Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nature Neuroscience 15:511–517.

Glassman EL (2005) A wavelet-like filter based on neuron action potentials for analysis of human scalp electroencephalographs. Institute of Electrical and Electronics Engineers Transactions of Biomedical Engineering 52:1851–1862.

Goodglass H (1993) Understanding aphasia. San Diego, CA, US: Academic Press.

Goutte C, Toft P, Rostrup E, Nielsen FÅ, Hansen LK (1999) On Clustering fMRI Time Series. NeuroImage 9:298–310.

Grami A (2019) Probability, Random Variables, Statistics, and Random Processes: Fundamentals & Applications. John Wiley & Sons.

Grancharov V, Samuelsson J, Kleijn WB (2005) Improved Kalman filtering for speech enhancement. In: Proceedings for Institute of Electrical and Electronics Engineers International Conference on Acoustics, Speech, and Signal Processing, 2005., pp I/1109-I/1112 Vol. 1. Philadelphia, PA: Institute of Electrical and Electronics Engineers.

Gruenwald J, Kapeller C, Kamada K, Scharinger J, Guger C (2017) Optimal bandpower estimation and tracking via Kalman filtering for real-time Brain-Computer Interfaces. In: 2017 8th International Institute of Electrical and Electronics Engineers Conference on Neural Engineering, pp 605–608. Shanghai, China: Institute of Electrical and Electronics Engineers.

Guenther FH (1994) A neural network model of speech acquisition and motor equivalent speech production. Biological Cybernetics 72:43–53.

Guenther FH (1995) Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. Psychology Review 102:594–621.

Guenther FH (2006) Cortical interactions underlying the production of speech sounds. Journal of Communication Disorders 39:350–365.

Guenther FH (2016) Neural Control of Speech. Cambridge, MA: MIT Press.

Guenther FH, Brumberg JS (2011) Brain-machine interfaces for real-time speech synthesis. Conference Proceedings of Institute of Electrical and Electronics Engineers Engineering Medicine and Biology Society 2011:5360–5363.

Guenther FH, Brumberg JS, Wright EJ, Nieto-Castanon A, Tourville JA, Panko M, Law R, Siebert SA, Bartels JL, Andreasen DS, Ehirim P, Mao H, Kennedy PR (2009) A Wireless Brain-Machine Interface for Real-Time Speech Synthesis. Public Library of Science One 4:e8218.

Guenther FH, Ghosh SS, Tourville JA (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. Brain and Language 96:280–301.

Guenther FH, Hampson M, Johnson D (1998) A theoretical investigation of reference frames for the planning of speech movements. Psychology Review 105:611–633.

Guenther FH, Tourville JA, Bohland JW (2015) Speech Production A2 - Toga, Arthur W. In: Brain Mapping, pp 435–444. Waltham, MA: Academic Press.

Gunduz A, Sanchez JC, Carney PR, Principe JC (2009) Mapping broadband electrocorticographic recordings to two-dimensional hand trajectories in humans Motor control features. Neural Networks 22:1257–1270.

Hagoort P (2013) MUC (Memory, Unification, Control) and beyond. Frontiers in Psychology 4:416.

Hagoort P, Indefrey P (2014) The neurobiology of language beyond single words. Annual Review of Neuroscience 37:347–362.

Halgren E, Dhond RP, Christensen N, Van Petten C, Marinkovic K, Lewine JD, Dale AM (2002) N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. Neuroimage 17:1101–1116.

Hamilton LS, Edwards E, Chang EF (2018) A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus. Current Biology 28:1860-1871.e4.

Haxby JV, Connolly AC, Guntupalli JS (2014) Decoding neural representational spaces using multivariate pattern analysis. Annual Review of Neuroscience 37:435–456.

Haynes K, Fearnhead P, A. Eckley I (2017) A computationally efficient nonparametric approach for changepoint detection. Statistics and Computing 27:1293–1305.

Henke WL (1966) Dynamic articulatory model of speech production using computer simulation. Massachusetts Institute of Technology, Cambridge, MA.

Herff C, Heger D, de Pesters A, Telaar D, Brunner P, Schalk G, Schultz T (2015) Brain-to-text: decoding spoken phrases from phone representations in the brain. Frontiers in Neuroscience 9:217.

Herman AB, Houde JF, Vinogradov S, Nagarajan SS (2013) Parsing the phonological loop: activation timing in the dorsal speech stream determines accuracy in speech reproduction. Journal of Neuroscience 33:5439–5453.

Hermes D, Miller KJ, Vansteensel MJ, Aarnoutse EJ, Leijten FSS, Ramsey NF (2012) Neurophysiologic correlates of fMRI in human motor cortex. Human Brain Mapping 33:1689–1699.

Hey T, Tansley S, Tolle K (2009) The Fourth Paradigm: Data-intensive Scientific Discovery. Redmond, WA: Microsoft Research.

Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69:407–422.

Hickok G, Poeppel D (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92:67–99.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nature Review of Neuroscience 8:393–402.

Honey CJ, Thesen T, Donner TH, Silbert LJ, Carlson CE, Devinsky O, Doyle WK, Rubin N, Heeger DJ, Hasson U (2012) Slow cortical dynamics and the accumulation of information over long timescales. Neuron 76:423–434.

Hope TMH, Prejawa S, Parker Jones ‘Ōiwi, Oberhuber M, Seghier ML, Green DW, Price CJ (2014) Dissecting the functional anatomy of auditory word repetition. Frontiers in Human Neuroscience 8:246.

Horwitz B, Friston KJ, Taylor JG (2000) Neural modeling and functional brain imaging: an overview. Neural Networks 13:829–846.

Hotelling H (1933) Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24:417–441.

Houde JF, Jordan MI (1998) Sensorimotor adaptation in speech production. Science 279:1213–1216.

Hsieh H-L, Shanechi MM (2018) Optimizing the learning rate for adaptive estimation of neural encoding models. Public Library or Science Computational Biology 14:e1006168.

Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat’s striate cortex. Journal of Physiology 148:574–591.

Hullett PW, Hamilton LS, Mesgarani N, Schreiner CE, Chang EF (2016) Human Superior Temporal Gyrus Organization of Spectrotemporal Modulation Tuning Derived from Speech Stimuli. Journal of Neuroscience 36:2014–2026.

Hupert JJ (1965) Modulation, noise, and spectral analysis: Applied to information transmission. Proceedings of the Institute of Electrical and Electronics Engineers 53:2171–2171.

Iacoboni M, Dapretto M (2006) The mirror neuron system and the consequences of its dysfunction. Nature Review of Neuroscience 7:942–951.

Indefrey P, Levelt WJM (2000) The neural correlates of language production. In: The New Cognitive Neurosciences, pp 845–865. Cambridge, MA: MIT Press.

Indefrey P, Levelt WJM (2004) The spatial and temporal signatures of word production components. Cognition 92:101–144.

Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17:825–841.

Jobert, Tismer, Poiseau, Schulz (1994) Wavelets-a new tool in sleep biosignal analysis. Journal of Sleep Research 3:223–232.

Jolliffe IT (2002) Principal Component Analysis, 2nd ed. New York, NY: Springer.

Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A - Mathematical Physical and Engineering Science 374:20150202.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiology Review 84:541–577.

Julier SJ, Uhlmann JK (2004) Unscented filtering and nonlinear estimation. Proceedings of the Institute of Electrical and Electronics Engineers 92:401–422.

Jurafsky D, Martin JH (2009) Speech and Language Processing, 2nd ed. Upper Saddle River, NJ: Pearson Education Inc.

Kalaska JF, Cohen DA, Hyde ML, Prud'homme M (1989) A comparison of movement direction-related versus load direction-related activity in primate motor cortex, using a two-dimensional reaching task. Journal of Neuroscience 9:2080–2102.

Kalman RE, Bucy RS (1961) New Results in Linear Filtering and Prediction Theory. Journal of Basic Engineering 83:95–108.

Kambara T, Brown EC, Jeong J-W, Ofen N, Nakai Y, Asano E (2017) Spatio-temporal dynamics of working memory maintenance and scanning of verbal information. Clinical Neurophysiology 128:882–891.

Karagiannidis GK, Lioumpas AS (2007) An Improved Approximation for the Gaussian Q-Function. Institute of Electrical and Electronics Engineers Communications Letters 11:644–646.

Kellis S, Miller K, Thomson K, Brown R, House P, Greger B (2010) Decoding spoken words using local field potentials recorded from the cortical surface. Journal of Neural Engineering 7:056007.

Keogh E, Ratanamahatana CA (2005) Exact Indexing of Dynamic Time Warping. Knowledge and Information Systems 7:358–386.

Klein A, Sauer T, Jedynak A, Skrandies W (2006) Conventional and wavelet coherence applied to sensory-evoked electrical brain activity. Institute of Electrical and Electronics Engineers Transactions of Biomedical Engineering 53:266–272.

Korzeniewska A, Franaszczuk PJ, Crainiceanu CM, Kuś R, Crone NE (2011) Dynamics of large-scale cortical interactions at high gamma frequencies during word production: event related causality (ERC) analysis of human electrocorticography (ECoG). Neuroimage 56:2218–2237.

Kubanek J, Brunner P, Gunduz A, Poeppel D, Schalk G (2013) The Tracking of Speech Envelope in the Human Cortex. Public Library of Science One 8:e53398.

Kubanek J, Schalk G (2015) NeuralAct: A Tool to Visualize Electrocortical (ECoG) Activity on a Three-Dimensional Model of the Cortex. Neuroinformatics 13:167–174.

Lachaux J-P, Fonlupt P, Kahane P, Minotti L, Hoffmann D, Bertrand O, Baciu M (2007) Relationship between task-related gamma oscillations and BOLD signal: new insights from combined fMRI and intracranial EEG. Human Brain Mapping 28:1368–1375.

Lavielle M (2005) Using penalized contrasts for the change-point problem. Signal Processing 85:1501–1510.

Leaver AM, Rauschecker JP (2010) Cortical Representation of Natural Complex Sounds: Effects of Acoustic Features and Auditory Object Category. Journal of Neuroscience 30:7604–7612.

Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401:788–791.

Lee H j, Roberts SJ (2008) On-line novelty detection using the Kalman filter and extreme value theory. In: 2008 19th International Conference on Pattern Recognition, pp 1–4. Tampa, FL: Institute of Electrical and Electronics Engineers.

Leonard MK, Baud MO, Sjerps MJ, Chang EF (2016) Perceptual restoration of masked speech in human cortex. Nature Communications 7:13619.

Leonard MK, Cai R, Babiak MC, Ren A, Chang EF (2019) The peri-Sylvian cortical networks underlying single word repetition revealed by electrocortical stimulation and direct neural recordings. Brain and Language 193:58–72.

Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011) Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. Journal of Neuroscience 31:2906–2915.

Leuthardt EC, Gaona C, Sharma M, Szrama N, Roland J, Freudenberg Z, Solis J, Breshears J, Schalk G (2011) Using the electrocorticographic speech network to control a brain-computer interface in humans. Journal of Neural Engineering 8:036004.

Leuthardt EC, Pei X-M, Breshears J, Gaona C, Sharma M, Freudenberg Z, Barbour D, Schalk G (2012) Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task. Frontiers of Human Neuroscience 6:99.

Li Z, O'Doherty JE, Hanson TL, Lebedev MA, Henriquez CS, Nicolelis MAL (2009) Unscented Kalman Filter for Brain-Machine Interfaces. Public Library of Science One 4:e6243.

Lichtheim L (1885) On Aphasia. Brain 7:433–484.

Lie OV, van Mierlo P (2017) Seizure-Onset Mapping Based on Time-Variant Multivariate Functional Connectivity Analysis of High-Dimensional Intracranial EEG: A Kalman Filter Approach. Brain Topography 30:46–59.

Lipton ZC (2016) The Mythos of Model Interpretability. Communications of the Associations for Computing Machinery 61:36–43.

Lloyd S (1982) Least squares quantization in PCM. Institute of Electrical and Electronics Engineers Transactions on Information Theory 28:129–137.

Lotte F, Brumberg JS, Brunner P, Gunduz A, Ritaccio AL, Guan C, Schalk G (2015) Electrocorticographic representations of segmental features in continuous speech. Frontiers in Human Neuroscience 9:97.

Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nature Neuroscience 9:1432–1438.

Magnotti JF, Beauchamp MS (2017) A Causal Inference Model Explains Perception of the McGurk Effect and Other Incongruent Audiovisual Speech. Public Library of Science Computational Biology 13:e1005229.

Majerus S (2013) Language repetition and short-term memory: an integrative framework. Frontiers of Human Neuroscience 7:357.

Makeig S (1993) Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroencephalography and Clinical Neurophysiology 86:283–293.

Mallat S (2009) A Wavelet Tour of Signal Processing: The Sparse Way, 3rd ed. Boston, MA: Academic Press.

Martin S, Brunner P, Holdgraf C, Heinze H-J, Crone NE, Rieger J, Schalk G, Knight RT, Pasley BN (2014) Decoding spectrotemporal features of overt and covert speech from the human cortex. Frontiers in Neuroengineering 7:14.

Martin S, Millán J del R, Knight RT, Pasley BN (2019) The use of intracranial recordings to decode human language: Challenges and opportunities. Brain and Language 193:73–83.

Mathe M, Nandyala SP, Kumar TK (2012) Speech enhancement using Kalman Filter for white, random and color noise. In: 2012 International Conference on Devices, Circuits and Systems, pp 195–198. Coimbatore, India: Institute of Electrical and Electronics Engineers.

McQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp 281–297. Berkeley, CA: University of California Press.

Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic Feature Encoding in Human Superior Temporal Gyrus. Science 343:1006–1010.

Miller KJ, Hermes D, Honey CJ, Hebb AO, Ramsey NF, Knight RT, Ojemann JG, Fetz EE (2012) Human motor cortical activity is selectively phase-entrained on underlying rhythms. Public Library of Science Computational Biology 8:e1002655.

Miller KJ, Sorensen LB, Ojemann JG, den Nijs M (2009) Power-law scaling in the brain surface electric potential. Public Library of Science Computational Biology 5:e1000609.

Mognon A, Jovicich J, Bruzzone L, Buiatti M (2011) ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology 48:229–240.

Moharramipour A, Mostame P, Hossein-Zadeh G-A, Wheless JW, Babajani-Feremi A (2018) Comparison of statistical tests in effective connectivity analysis of ECoG data. Journal of Neuroscience Methods 308:317–329.

Moore CJ, Price CJ (1999) Three Distinct Ventral Occipitotemporal Regions for Reading and Object Naming. Neuroimage 10:181–192.

Moritz-Gasser S, Duffau H (2013) The anatomo-functional connectivity of word repetition: insights provided by awake brain tumor surgery. Frontiers in Human Neuroscience 7:405.

Moses DA, Mesgarani N, Leonard MK, Chang EF (2016) Neural speech recognition: continuous phoneme decoding using spatiotemporal representations of human cortical activity. Journal of Neural Engineering 13:056004.

Moussakhani B, Flåm JT, Ramstad T, Balasingham I (2014) On change detection in a Kalman filter based tracking problem. Signal Processing 105:268–276.

Mugler EM, Patton JL, Flint RD, Wright ZA, Schuele SU, Rosenow J, Shih JJ, Krusienski DJ, Slutzky MW (2014) Direct classification of all American English

phonemes using signals from functional speech motor cortex. Journal of Neural Engineering 11:035015.

Mugler EM, Tate MC, Livescu K, Templer JW, Goldrick MA, Slutzky MW (2018) Differential Representation of Articulatory Gestures and Phonemes in Precentral and Inferior Frontal Gyri. Journal of Neuroscience 38:9803–9813.

Nourski KV, Steinschneider M, McMurray B, Kovach CK, Oya H, Kawasaki H, Howard MA (2014) Functional organization of human auditory cortex: investigation of response latencies through direct recordings. Neuroimage 101:598–609.

Nourski KV, Steinschneider M, Rhone AE, Iii H, A M (2017) Intracranial Electrophysiology of Auditory Selective Attention Associated with Speech Classification Tasks. Frontiers in Human Neuroscience 10:691.

Ojemann G, Ojemann J, Lettich E, Berger M (1989) Cortical language localization in left, dominant hemisphere. An electrical stimulation mapping investigation in 117 patients. Journal of Neurosurgery 71:316–326.

Ojemann GA, Ramsey NF, Ojemann J (2013) Relation between functional magnetic resonance imaging (fMRI) and single neuron, local field potential (LFP) and electrocorticography (ECoG) activity in human cortex. Frontiers in Human Neuroscience 7:34.

O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Research 34:171–175.

Ozker M, Schepers IM, Magnotti JF, Yoshor D, Beauchamp MS (2017) A Double Dissociation between Anterior and Posterior Superior Temporal Gyrus for Processing Audiovisual Speech Demonstrated by Electrocorticography. Journal of Cognitive Neuroscience 29:1044–1060.

Page ES (1963) Controlling the Standard Deviation by Cusums and Warning Lines. Technometrics 5:307–315.

Parker Jones 'Ōiwi, Prejawa S, Hope T, Oberhuber M, Seghier ML, Leff AP, Green DW, Price CJ (2014) Sensory-to-motor integration during auditory repetition: a combined fMRI and lesion study. Frontiers in Human Neuroscience 8:24.

Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF (2012) Reconstructing Speech from Human Auditory Cortex. Public Library of Science Biology 10:e1001251.

Paulesu E, Frith CD, Frackowiak RS (1993) The neural correlates of the verbal component of working memory. Nature 362:342–345.

Pearson K (1901) On Lines and Planes of Closest Fit to Systems of Points in Space. Philosophical Magazine 2:559–572.

Pei X, Barbour DL, Leuthardt EC, Schalk G (2011a) Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. Journal of Neural Engineering 8:046028.

Pei X, Leuthardt EC, Gaona CM, Brunner P, Wolpaw JR, Schalk G (2011b) Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. Neuroimage 54:2960–2972.

Penfield W, Boldrey E (1937) Somatic Motor and Sensory Representation in the Cerebral cortex of Man as Studied by Electrical Stimulation. Brain 60:389–443.

Penfield W, Roberts L (1959) Speech and Brain Mechanisms. Princeton, NJ: Princeton University Press.

Pfurtscheller G, Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. Clinical Neurophysiology 110:1842–1857.

Potes C, Gunduz A, Brunner P, Schalk G (2012) Dynamics of electrocorticographic (ECoG) activity in human temporal and frontal cortical areas during music listening. Neuroimage 61:841–848.

Potworowski J, Jakuczun W, Lęski S, Wójcik D (2012) Kernel current source density method. Neural Computation 24:541–575.

Price CJ, Wise RJ, Warburton EA, Moore CJ, Howard D, Patterson K, Frackowiak RS, Friston KJ (1996) Hearing and saying. The functional neuro-anatomy of auditory word processing. Brain 119:919–931.

Rao VR, Leonard MK, Kleen JK, Lucas BA, Mirro EA, Chang EF (2017) Chronic ambulatory electrocorticography from human speech cortex. Neuroimage 153:273–282.

Rauschecker JP (1998) Cortical processing of complex sounds. Current Opinions of Neurobiology 8:516–521.

Ray S, Crone NE, Niebur E, Franaszczuk PJ, Hsiao SS (2008) Neural correlates of high-gamma oscillations (60-200 Hz) in macaque local field potentials and their potential implications in electrocorticography. Journal of Neuroscience 28:11526–11536.

Ray S, Maunsell JHR (2011) Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. Public Library of Science Biology 9:e1000610.

Rosso OA, Blanco S, Rabinowicz A (2003) Wavelet analysis of generalized tonic-clonic epileptic seizures. Signal Processing 83:1275–1289.

Rottschy C, Langner R, Dogan I, Reetz K, Laird AR, Schulz JB, Fox PT, Eickhoff SB (2012) Modelling neural correlates of working memory: a coordinate-based meta-analysis. Neuroimage 60:830–846.

Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20:53–65.

Rubin P, Baer T, Mermelstein P (1981) An articulatory synthesizer for perceptual research. The Journal of the Acoustical Society of America 70:321–328.

Rubino (2006) Propagating waves mediate information transfer in the motor cortex. Nature Neuroscience 9:1549–1557.

Saltzman EL, Munhall KG (1989) A dynamical approach to gestural patterning in speech production. Ecological Psychology 1:333–382.

Samar VJ, Bopardikar A, Rao R, Swartz K (1999) Wavelet Analysis of Neuroelectric Waveforms: A Conceptual Tutorial. Brain and Language 66:7–60.

Schönwiesner M, Zatorre RJ (2009) Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proceedings of the National Academy of Sciences of the United States of America 106:14611–14616.

Senhadji L, Wendling F (2002) Epileptic transient detection: wavelets and time-frequency approaches. Neurophysiologie Clinique 32:175–192.

Serrà J, Arcos JL (2014) An Empirical Evaluation of Similarity Measures for Time Series Classification. Knowledge-Based Systems 67:305–314.

Severo M, Gama J (2006) Change Detection with Kalman Filter and CUSUM. In: DS'06 Proceedings of the 9th international conference on Discovery Science, pp 243–254. Barcelona, Spain: Springer-Verlag.

Sharpee T, Rust NC, Bialek W (2004) Analyzing neural responses to natural signals: maximally informative dimensions. Neural Computation 16:223–250.

Simon MK (2007) Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers and Scientists. New York, NY: Springer.

Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. Journal of Acoustic Society of America 114:3394–3411.

Slepian D (1978) Prolate spheroidal wave functions, fourier analysis, and uncertainty #x2014; V: the discrete case. The Bell System Technical Journal 57:1371–1430.

Song L, Kolar M, Xing EP (2009) Time-Varying Dynamic Bayesian Networks. In: Advances in Neural Information Processing Systems 22, pp 1732–1740. Vancouver, British Columbia, Canada: Curran Associates, Inc.

Soule A, Salamatian K, Taft N (2005) Combining filtering and statistical methods for anomaly detection. In: 5th Association of Computing Machinery Signals and Communication conference on Internet measurement, pp 31. Berkeley, CA: Advanced Computing Systems Association.

Steinschneider M, Nourski KV, Kawasaki H, Oya H, Brugge JF, Howard MA (2011) Intracranial study of speech-elicited activity on the human posterolateral superior temporal gyrus. Cerebral Cortex 21:2332–2347.

Stephen EP (2015) Characterizing dynamically evolving functional networks in humans with application to speech. Boston University, Boston, MA.

Stephen EP, Lepage KQ, Eden UT, Brunner P, Schalk G, Brumberg JS, Guenther FH, Kramer MA (2014) Assessing dynamics, spatial scale, and uncertainty in task-related brain network analyses. Frontiers in Computational Neuroscience 8:31.

Stevens KN (2000) Acoustic Phonetics. Cambridge, MA: MIT Press.

Sturm I, Blankertz B, Potes C, Schalk G, Curio G (2014) ECoG high gamma activity reveals distinct cortical representations of lyrics passages, harmonic and timbre-related changes in a rock song. Frontiers in Human Neuroscience 8:798.

Takai O, Brown S, Liotti M (2010) Representation of the speech effectors in the human motor cortex: somatotopy or overlap? Brain and Language 113:39–44.

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL (2001) Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network 12:289–316.

Thomson DJ (1982) Spectrum estimation and harmonic analysis. Proceedings of the Institute of Electrical and Electronics Engineers 70:1055–1096.

Thorndike RL (1953) Who belongs in the family? Psychometrika 18:267–276.

Tolhurst DJ, Movshon JA, Dean AF (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Research 23:775–785.

Tourville JA, Guenther FH (2011) The DIVA model: A neural theory of speech acquisition and production. Language and Cognitive Process 26:952–981.

Towle VL, Yoon H-A, Castelle M, Edgar JC, Biassou NM, Frim DM, Spire J-P, Kohrman MH (2008) ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. Brain 131:2013–2027.

Turkeltaub PE, Eden GF, Jones KM, Zeffiro TA (2002) Meta-Analysis of the Functional Neuroanatomy of Single-Word Reading: Method and Validation. Neuroimage 16:765–780.

Turner R (2016) A Model Explanation System: Latest Updates and Extensions. In: Institute of Electrical and Electronics Engineers 26th International Workshop on Machine Learning for Signal Processing (MLSP), pp 1–6. Vietri sul Mare, Italy: Institute of Electrical and Electronics Engineers.

Tuyisenge V, Trebaul L, Bhattacharjee M, Chanteloup-Forêt B, Saubat-Guigui C, Mîndruţă I, Rheims S, Maillard L, Kahane P, Taussig D, David O (2018) Automatic bad channel detection in intracranial electroencephalographic recordings using ensemble machine learning. Clinical Neurophysiology 129:548–554.

Vallar G, Di Betta AM, Silveri MC (1997) The phonological short-term store-rehearsal system: patterns of impairment and neural correlates. Neuropsychologia 35:795–812.

Varoquaux G, Thirion B (2014) How machine learning is shaping cognitive neuroimaging. Gigascience 3:28.

Vigneau M, Jobard G, Mazoyer B, Tzourio-Mazoyer N (2005) Word and non-word reading: what role for the Visual Word Form Area? Neuroimage 27:694–705.

Wang X (2010) Neurophysiological and computational principles of cortical rhythms in cognition. Physiology Review 90:1195–1268.

Warburton E, Wise RJ, Price CJ, Weiller C, Hadar U, Ramsay S, Frackowiak RS (1996) Noun and verb retrieval by normal subjects. Studies with PET. Brain 119:159–179.

Warren Liao T (2005) Clustering of time series data—a survey. Pattern Recognition 38:1857–1874.

Wernicke C (1874) Der aphasiche symptomencomplex. Breslau: Cohn and Wigert.

Wessinger CM, Buonocore MH, Kussmaul CL, Mangun GR (1997) Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. Human Brain Mapping 5:18–25.

Xia Y, Wang J (2015) Low-dimensional recurrent neural network-based Kalman filter for speech enhancement. Neural Networks 67:131–139.

Zhang D, Gong E, Wu W, Lin J, Zhou W, Hong B (2012) Spoken sentences decoding based on intracranial high gamma response using dynamic time warping. In: 2012 Annual International Conference of the Institute of Electrical and Electronics Engineers Engineering in Medicine and Biology Society, pp 3292–3295.

Zhou W, Shu H (2017) A meta-analysis of functional magnetic resonance imaging studies of eye movements and visual word reading. Brain Behavior 7:e00683.

# CURRICULUM VITAE